DOCUMENT RESUME

ED 280 860 TM 870 186

TITLE	Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report Synopsis, 1984 Fiscal Year.
INSTITUTION	Human Resources Research Organization, Alexandria, Va.
SPONS AGENCY	Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
REPORT NO	ARI-RR-1393
PUB DATE	Jul 85
CONTRACT	MDA-903-82-C-0531
NOTE	79p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Computer Assisted Testing; Criterion Referenced Tests; Databases; Job Performance; *Job Placement; *Military Personnel; *Occupational Tests; Performance Tests; *Predictive Measurement; Research and Development; *Test Construction; Test Theory; *Test Validity
IDENTIFIERS	Armed Services Vocational Aptitude Battery; *Army

ABSTRACT

This report describes technical plans and progress during the second year (Fiscal Year 1984) of work on the U.S. Army's Project A: "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel." Project A is a long-range research program developed by the Army Research Institute for the Behavioral and Social Sciences. The goal is to establish a computerized personnel allocation system to match available personnel resources with Army manpower requirements based on biographical, psychological, and performance measures and a firm quantification of their interrelationships. The second year's activities focused on the following: (1) evaluation of the validity and fairness of existing prediction and criterion measures, such as the Armed Services Vocational Aptitude Battery; (2) development of criterion measures that reflect all major factors of job performance for first-tour enlisted personnel; (3) development and pilot testing of predictor measures called the Pilot Trial Battery; and (4) validation research based on the expansion of the longitudinal Research Database. This validation research led to proposed improvements in the Army's existing procedures for selecting and classifying new recruits. The status and the future directions of army selection and classification research are also discussed. Abstracts of 26 reports, articles, and papers referenced in the test are appended. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED280860

Research Report 1393

Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report Synopsis, 1984 Fiscal Year

Human Resources Research Organization
American Institutes for Research
Personnel Decisions Research Institute
Army Research Institute

Selection and Classification Technical Area

Manpower and Personnel Research Laboratory

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U. S. Army

Research Institute for the Behavioral and Social Sciences

July 1985

Approved for public release; distribution unlimited.

TM 870186

BEST COPY AVAILABLE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Report 1393	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: ANNUAL REPORT SYNOPSIS, 1984 FISCAL YEAR		5. TYPE OF REPORT & PERIOD COVERED Final/Annual Report 1 Oct 83 - 30 Sep 84 6. PERFORMING ORG. REPORT NUMBER ---
7. AUTHOR(s) Human Resources Research Organization American Institutes for Research Personnel Decisions Research Institute Army Research Institute		8. CONTRACT OR GRANT NUMBER(s) MDA 903-82-C-0531
9. PERFORMING ORGANIZATION NAME AND ADDRESS Human Resources Research Organization 1100 So. Washington Street Alexandria, VA 22314		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263731A792
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Ave., Alexandria, VA 22333-5600		12. REPORT DATE Jul 85 13. NUMBER OF PAGES 79
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) ---		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

---

18. SUPPLEMENTARY NOTES

The Army Research Institute technical point of contact is Dr. Newell K. Eaton. His telephone number is (202) 274-8275.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Performance measures	Selection	Army-wide measures
Predictor measures	Classification	Aptitude area composites
Validation	Criterion measures	Longitudinal data base
Predictors	Job knowledge tests	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the research performed during the second year of a project to develop a complete personnel system for selecting and classifying all entry-level Army enlisted personnel. In general, the second year's activities have emphasized the evaluation of the validity and fairness of existing prediction and criterion measures, and the development and initial testing of improved methods of predicting and measuring performance. This synopsis of the year's work is supplemented by ARI Technical Report 660, which documents in the

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (Continued) ARI Research Report 1393
context of the annual report a variety of technical reports and papers associated with the project, and by ARI Research Note 85-14, which provides supplementary appendix material for several papers included in Technical Report 660.

Research Report 1393

Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report Synopsis, 1984 Fiscal Year

Human Resources Research Organization
American Institutes for Research
Personnel Decisions Research Institute
Army Research Institute

Selection and Classification Technical Area
Newell K. Eaton, Chief

Manpower and Personnel Research Laboratory
Joyce L. Shields, Director

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

July 1985

Army Project Number
2Q263731A792

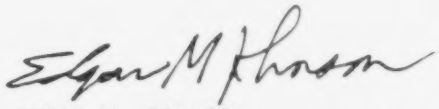
Manpower and Personnel

Approved for public release; distribution unlimited.

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

This document is a synopsis of the second year of research on the Army's current, large-scale manpower and personnel effort for improving the selection, classification, and utilization of Army enlisted personnel. The thrust for the project came from the practical, professional, and legal need to validate the Armed Services Vocational Aptitude Battery (ASVAB--the current U.S. military selection/classification test battery) and other selection variables as predictors of training and performance. The portion of the effort described herein is devoted to the development and validation of Army selection and classification measures and is referred to as "Project A." A second component, the development of a prototype Computerized Personnel Allocation System, is referred to as "Project B." Together, these Army Research Institute research efforts, with their in-house and contract components, compose a landmark program to develop a state-of-the-art, empirically validated system for personnel selection, classification, and allocation.



EDGAR M. JOHNSON
Technical Director

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: ANNUAL REPORT SYNOPSIS, 1984 FISCAL YEAR

PREFACE

This is a synopsis of the second year of research conducted on Project A, "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel." The project addresses the 675,000-person enlisted personnel system of the U.S. Army, with several hundred different occupations, from infantryman to typist to medic to mechanic. The goal is a computerized personnel allocation system to match available personnel resources with Army manpower requirements, based on biographical, psychological, and performance measures and a firm quantification of their interrelationships.

The research is being accomplished by one team of researchers addressing predictor and performance measures and their interrelationships, and by a second team using those measures to develop an allocation system (efforts in these areas have been termed "Project A" and "Project B," respectively).

The planning for this research was initiated by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) in 1980. As in-house resources were evaluated, it became apparent that the massive scope of the effort would be best met by a combination of the talents of research scientists and managers from ARI as well as contract research organizations. In 1981 ARI in-house scientists set to work developing the basic research requirements for the effort.

In 1982 a consortium led by the Human Resources Research Organization (HumRRO), and including the American Institutes for Research (AIR) and the Personnel Decisions Research Institute (PDRI), was selected by ARI as the contract organization offering the most innovative and creative approaches to meet the objectives of Project A. Scientists from ARI and the consortium, together with a multitude of advisors, developed a research plan to guide the project (U.S. Army Research Institute Research Report 1332, May 1983). The present report is a synopsis of the second year of research conducted according to that plan, with elaborations and changes outlined in the following sections.

Each section of this synopsis describes the efforts of many scientists in the consortium and ARI. Papers and reports based on their efforts are abstracted in the last pages of the synopsis, and published in the second Project A annual report (Eaton, Goer, Harris, and Zook, ARI Technical Report 660, October, 1984 unless they have been previously published separately. Principal authors of the sections of this synopsis are noted below:

I. The "Project A" Research Program
Newell K. Eaton, Marvin H. Goer, and Lola M. Zook

II. School and Job Performance Measurement
John P. Campbell

III. Predictor Measurement
Norman G. Peterson

IV. Validation
Paul G. Rossmeissl and Lauress L. Wise

V. Status and Future Directions of Army Selection and Classification Research
John P. Campbell and Newell K. Eaton

The major challenge of the third year of the project is the concurrent validation of the measures with 12,000 soldiers. The project will continue to evolve through continued discourse among the Army's senior leadership, representatives of the Department of the Defense and the Joint Services, the scientific community, and the ARI and contractor scientists. The aims are to provide the Army with a greatly improved, empirically based personnel system responsive to the needs of the service, while considering the unique abilities, interests, and desires of individual soldiers, and to enhance substantially the scientific knowledge in applied personnel selection and classification research.

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: ANNUAL REPORT SYNOPSIS, 1984 FISCAL YEAR

CONTENTS

		Page
I.	THE "PROJECT A' RESEARCH PROGRAM	1
	"Project A" Research Design	1
	Overview of Second-Year Progress	4
	Project Administration	6
	Documentation	9
II.	SCHOOL AND JOB PERFORMANCE MEASUREMENT	10
	Objectives for FY84	10
	Construction of Initial Measures	11
	Batch A(X) Field Tests	14
	Confirmatory Analysis: A Beginning	17
	Discussion and Conclusions	17
	Potential Applications of FY84 Criterion Development Products	21
	Documentation	22
III.	PREDICTOR MEASUREMENT	23
	Predictor Development	23
	Pilot Trial Battery	26
	Summation	29
	Documentation	29
IV.	VALIDATION	30
	Growth of the LRDB	30
	ASVAB Area Composite Validation	32
	LRDB Support Activities	36
	Documentation	36
V.	STATUS AND FUTURE DIRECTIONS OF ARMY SELECTION AND CLASSIFICATION RESEARCH	38
REFERENCES		41
APPENDIX A.	ABSTRACTS FOR REPORTS, ARTICLES, AND PAPERS REFERENCED IN THE TEXT, FOR THE 1984 FISCAL YEAR	A-1

Page

LIST OF TABLES

Table 1. "Batch A" field test samples 16

2. FY83/84 soldiers with preliminary battery and training data . . 32

LIST OF FIGURES

Figure 1. The research flow 2

2. Project A MOS 3

3. Governance Advisory Group 8

4. Project organization 8

5. Three alternative scenarios for SME judgments of task and item importance 20

6. Flow chart of predictor measure development activities on Project A 24

7. Cognitive/perceptual/psychomotor measures in the pilot trial battery 27

8. Non-cognitive measures in the pilot trial battery: The Army Vocational Interest and Career Examination (AVOICE) and the Assessment of Background and Life Experiences (ABLE) 28

9. Validity analyses sample sizes 33

10. Predictive validities systems for nine and four composites . . 34

11. A comparison of current and alternative composites 35

I. THE "PROJECT A" RESEARCH PROGRAM

The purpose of this annual report is to describe technical plans and progress during the second year (Fiscal Year 1984) of work on the U.S. Army's Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel. Project A is a comprehensive, long-range research program developed by the Army Research Institute for the Behavioral and Social Sciences (ARI). Our goal is a computerized personnel allocation system to match available personnel resources with Army manpower requirements, based on biographical, psychological, and performance measures and a firm quantification of their interrelationships.

The 9-year project employs 40-50 researchers in a variety of specialties of industrial and organizational psychology, operations research, management science, and computer science. It addresses the 675,000-person enlisted personnel system of the U.S. Army, which encompasses several hundred different occupations, from infantryman to typist to medic to mechanic.

A major focus of the project is the development of new predictor and criterion measures to expand the dimensions and improve the accuracy of measurement of the respective predictor and criterion space. There appears to be a heavy general-ability (Spearman's "G") loading in both the paper-and-pencil Armed Services Vocational Aptitude Battery (ASVAB) and the Skill Qualification Tests (SQT) currently used by the Army. This research is designed to provide measures that more completely encompass the full range of potential performance predictors and to provide criterion measures that more adequately represent actual job performance. In each military occupational specialty (MOS) the most valid composite of predictors will be used as selection/classification factors to provide the best person-job match for overall soldier performance.

"Project A" Research Design

The Project A research design incorporates three iterations of data collection and analysis to provide timely and responsive results during the course of the effort. It also permits the correction of errors and the exploitation of opportunities. A schematic of the design is shown in Figure 1.

In the first iteration, file data from fiscal year (FY) 1981 and 1982 accessions were evaluated to verify the empirical linkage between existing ASVAB scores and subsequent training and first-tour knowledge test performance.

In the second iteration, a predictive-concurrent design is being executed with FY83/84 accessions. Several thousand soldiers in four occupations have been tested at entry on a preliminary battery of spatial, perceptual, temperament/personality, interest, and biodata measures. These soldiers' data were entered into a longitudinal research data base (LRDB) containing operational ASVAB and other enlistment measures on all FY83-84 accessions.

About 600 soldiers in each of these four MOS, and in each of an additional 15 MOS, will be tested in FY85. A revised test battery, including computer-administered perceptual and psychomotor predictor instruments, will be concurrently administered with a set of job-specific and general performance indices based on knowledge, hands-on (for half the MOS), and rating measures. About a hundred soldiers in each MOS will be retested after three years, during their second Army tour.

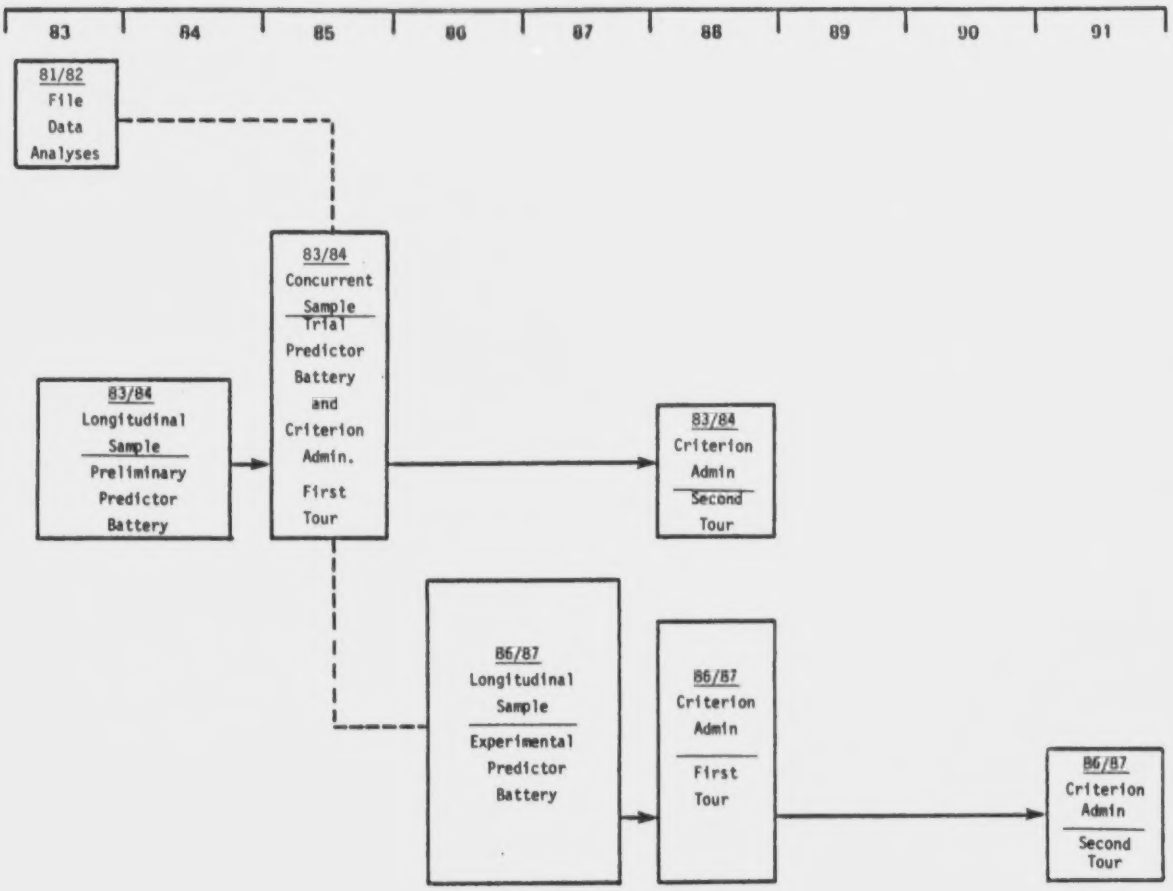


Figure 1. The research flow.

The 19 MOS chosen for testing comprise a specially selected representative sample of the 250 entry-level MOS. They are shown in Figure 2 (Batch A, B, and Z groupings, explained later, are indicated). The MOS selection was based on an initial clustering of MOS, derived from rated similarities of job content. These 19 MOS account for about 45 percent of Army accessions. Sample sizes are sufficient to empirically evaluate race and sex fairness in most MOS.

BATCH A

MOS	Title	FY83 Accessions
13B	Cannon Crewman	6,431
64C	Motor Transport Oper	4,282
71L	Admin Specialist	5,219
95B	Military Police	5,873

BATCH B

MOS	Title	FY83 Accessions
05C	Radio TT Oper	1,815
11B	Infantryman	15,904
19E/K	Tank Crewman	3,935
63B	Vehicle & Generator Mech	4,807
91B	Medical Care Specialist	4,681

BATCH Z

MOS	Title	FY83 Accessions
12B	Combat Engineer	1,554
16S	MANPADS Crewman	624
27E	Tow/Dragon Rpr	254
51B	Carpentry/Masonry Spec	183
54E	Chemical Operations Spec	1,302
56B	Ammunition Spec	571
67N	Utility Helicopter Rpr	621
75W	Petroleum Supply Spec	1,205
76Y	Unit Supply Spec	3,651
94B	Food Service Spec	5,375
	TOTAL	134,696

Figure 2. Project A MOS

In the third iteration, all of the measures, refined by the experiences of the first and second iteration, will be collected sequentially in a true predictive validity design. About 50,000 soldiers across about 20 MOS will be included in the FY86-87 predictor battery administration. After losses from all factors, about 3,500 will be included in second-tour performance measurement in FY91.

The detailed research plan is described in ARI Research Report 1332, May 1983. The initial plan had been expanded and refined during the first few months of work on the project, which began in October 1982.

Overview of Second-Year Progress

During the second year of work on Project A, major gains have been made in development of performance measures and prediction tests, evaluation of the validity and the race/sex fairness of the ASVAB, and development of utility measures. The work is described in the present report and in a companion report, ARI Technical Report 660, "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year" (October 1984). The latter report includes various technical documents that have been prepared during the year to report on specialized aspects of the research program. (These reports are listed in the present volume in the relevant sections and abstracts are provided in Appendix A.) The Technical Report is supplemented by ARI Research Note 85-14, which supplies appendix material (research instruments and analyses) for two papers contained in the Technical Report.

Plans for the project as a whole and activities during the first year were described in the annual report for the 1983 fiscal year, ARI Research Report 1347, and the technical appendix to that report, ARI Research Note 83-37, both published in October 1983.

Performance Measurement. The research effort on performance measures has developed nicely. We have developed an extensive task inventory for the first 19 key MOS, based on Soldier's Manuals, Army Occupational Survey Program, and data from subject matter experts. Efforts have been made to level the generality of task descriptions, and to determine the variability of performance, importance, and frequency of each task. This detailed analysis provides a firm basis for both knowledge and hands-on task sampling.

Field tests have been conducted with 150 soldiers in each of the first four MOS (Batch A): clerk-typist (71L), military police (95B), driver (64C), and artillery crewman (13B). Field tests for five more MOS (Batch B) will be completed in the spring of 1985. Tests on 30 tasks representing each MOS are administered in a paper-and-pencil format; 15 are also administered in a hands-on mode. Ratings from peers and supervisors are also obtained on the soldier's ability to perform these tasks. Additionally, measurements of organizational variables and knowledge of information presented during training, as well as ratings of general soldiering behaviors, are collected during the field tests.

Information obtained from the field tests, and during the FY85 tests, will inform our decisions on the most efficient manner in which to construct comprehensive job performance measures. Preliminary information, from two of the first four MOS field tested, indicates relatively high internal consistency within measurement method, but relative independence between methods.

We expect that the results of the field tests and FY85 tests will provide strong evidence that will affect criterion development. Questions of "ultimate" criteria, and the parameters determining the relationships between hands-on tests, job knowledge, and peer or supervisory ratings, will be addressed. Because complete data will be available in nine diverse MOS (Batches A and B), and partial data in 10 more (Batch Z), we expect to obtain relatively comprehensive answers to these questions.

Another question is how to determine minimum performance standards. We are beginning by presenting our quantitative performance distributions in proponent workshops. Both trainers and leaders in operational units will see how soldiers in their occupations performed or were rated on all the measures, and how the measures are intercorrelated. Through their individual judgments and consensual feedback procedures, we will attempt to elicit minimum performance standards for approval by Army policymakers. These will inform policymakers' decisions on acceptable predictor scores for entry into MOS.

Predictor Measurement. In our predictor development the taxonomy of human abilities presented by Peterson and Bownas (1982) was used as a starting point. Based on an exhaustive literature review followed by analyses of expert judgments of predictor-criterion validity coefficients, a predictor-by-performance factors matrix was created. Twenty-five predictor constructs are currently being considered for administration to the FY83/84 cohort in FY85. Four of the predictor constructs are measured by the current ASVAB. Twelve more were measured in the predictive design portion of the second design iteration, for accessions in four MOS. In addition, field tests have been completed on seven microprocessor-based cognitive, perceptual and psychomotor constructs. Of significant interest is the relative independence of these measures. We appear to be well on the way to extending the predictor space beyond "G".

Validation. A longitudinal research data base, containing data on Army applicants beginning in FY81 and continuing through the present time, is one of our major accomplishments. After countless hours of file cleaning, sorting, and patching, we have records on more than 600,000 applicants and more than 300,000 accessions. Predictor information consists of operational accessions records data: ASVAB, the Military Applicant Profile (MAP) for non-graduates, and some other biodata. Performance data consist of end-of-course training data reported by the schools (FY81 only), SQTs, and data from the Enlisted Master File: attrition, promotion, disciplinary actions, awards, etc.

The first iteration of the data collection specified in the research design is complete. This step included the analysis of the validity of the current ASVAB as a predictor of MOS training and first-tour SQT performance. The results were based on a sample in excess of 60,000 soldiers. They demonstrated the validity of the nine operational ASVAB composites, with a median validity of .48 for training and SQT combined.

Further, the results showed that a change in the composition of two composites, CL (clerical) and SC (surveillance and communication), produced an increase in predictive validity. The Army operationalized these new composites beginning in October 1984, an action that will improve the prediction of performance of 20,000 soldiers entering each year.

The utility of any selection or classification effort is an important issue, and there has been a significant rebirth of interest in this area in the last five years. On the basis of an estimation technique developed by Schmidt, Hunter, McKenzie, and Muldrow (1979), the dollar value of the Army's change in the CL and SC composites was estimated to be $5,000,000

per year. The effort toward better ways to evaluate the utility of selection and classification efforts provided both an extension to the Schmidt et al. method that appears to be more appropriate in military settings, and an entirely new method. Substantial progress is also being made in a utility effort designed to evaluate the relative worth of various levels of performance within and between MOS; the pilot efforts have used the 50th percentile infantryman as a standard.

Project Administration

The overall administration and structure of the Project A research effort continued without change in FY84. For administrative purposes, Project A is organized into major tasks (Task 1, Validation; Task 2, Developing Predictors of Job Performance; Task 3, Measurement of School/Training Success; Task 4, Assessment of Army-wide Performance; Task 5, Develop MOS-Specific Performance Measures; Task 6, Management). The research efforts under the various tasks are interrelated and integrated through continuous oversight by Task 6 in-house and contractor staffs as well as the regular programs of Interim Progress Review (IPR) meetings and discussions.

Contract Amendment. ARI Research Report 1332, "Improving the Selection, Classification and Utilization of Army Enlisted Personnel--Project A: Research Plan" (May 1983), specified a number of changes to the original scope of work described in the RFP. These changes required that an amendment to the contract be formulated and approved to bring it into conformance with the Project A Research Plan.

The amendment provides for a shift in focus to future cohorts (from the FY81/82 and FY84/85 cohorts to the FY83/84 and FY86/87 cohorts. It also specifies the additional work entailed in:

- Acquiring school data on the FY83/84 cohort for predictor and criterion development.
- Conducting validity analyses of FY81/82 cohort data in support of mandated Aptitude Area Composite recommendations.
- Conducting job and task analyses to support new "cluster" constructs, and identifying the focal MOS.
- Preparing detailed analyses and justification to support the sampling strategy (and the resultant Troop Support Requests).
- Accomplishing a "Preliminary Battery" identification and test phase in the predictor development and test research program.
- Acquiring, using, and maintaining psychomotor/perceptual test equipment in the new predictor Trial and Experimental Battery research and development program.
- Expanding the utility research program to include the requirements for development of "monetization" metrics.

- Extending the research schedule through 1991 to retain the objective of analyzing second-term validity data on the second (FY86/87) main cohort.

In December 1983, ARI informed the consortium managers that funding plans for the second year of contract performance would have to conform to funding limitations and that the research program activities would have to be adjusted accordingly. Concurrent with accommodating to FY84 fund limitations, it was determined that the estimate of resources required for scientific quality assurance and control, interim product development and exploitation, an expanded program of communications and reporting, and maintenance of intertask coordination and interface was insufficient for a program of this scope and complexity. Accordingly, the amendment to the contract provided resources for meeting these new requirements and constraints.

An amendment proposal for the contract was provided to ARI 20 April 1984 and subjected to an intensive review and evaluation process. On 28 September 1984 the amendment was approved and was incorporated into the contract.

Psychomotor/Perceptual Test Equipment. Included in the changes noted above was a requirement for an extensive investigation of psychomotor/perceptual constructs to meet the objective of researching the broadest spectrum of potential predictors, thereby providing a better possibility of improving on the ASVAB. Implementing this decision required the acquisition, use, and maintenance of psychomotor/perceptual equipment for development work and the subsequent major data collections planned for the FY83/84 and FY86/87 main cohorts.

During FY84, all of the procedures and requirements of AR 18-1, governing the acquisition of computers, were fully complied with; this included the development and provision of a satisfactory Mission Element Need Statement (MENS), an Acquisition Plan, and an Economic Analysis supporting and justifying the requirement for the psychomotor/perceptual testing equipment. These documents were reviewed by the cognizant Army organizations, and the acquisition was approved 2 August 1984 by the Assistant Secretary of the Army (Financial Management).

Personnel Changes. During the course of the second year's work a number of personnel changes were effected in the Governance Advisory Group. BG W. C. Knudson (Office of the Deputy Chief of Staff for Operations and Plans) and BG Frederick M. Franks, Jr. (USAREUR) were designated as U.S. Army Advisors. In addition, Dr. W. S. Sellman replaced Dr. G. T. Sicilia as the DoD Interservice Advisor. These changes are reflected in Figure 3.

There were also changes in assignments for the ARI Task Monitors and Consortium Task Leaders and other key personnel. The assignments for these monitor/leader positions at the end of FY84 are reflected in Figure 4. To help in providing the best advice and evaluation of task activities, members of the Scientific Advisory Group have agreed to place special emphasis on specific Tasks, and monitor Task progress at semiannual in-process reviews. Dr. Linn is aligned with Task 1, Drs. Humphreys and Uhlaner with Task 2, Dr. Hakel with Task 3, Dr. Bobko with Task 4, and Drs. Cook and Tenopyr with Task 5.

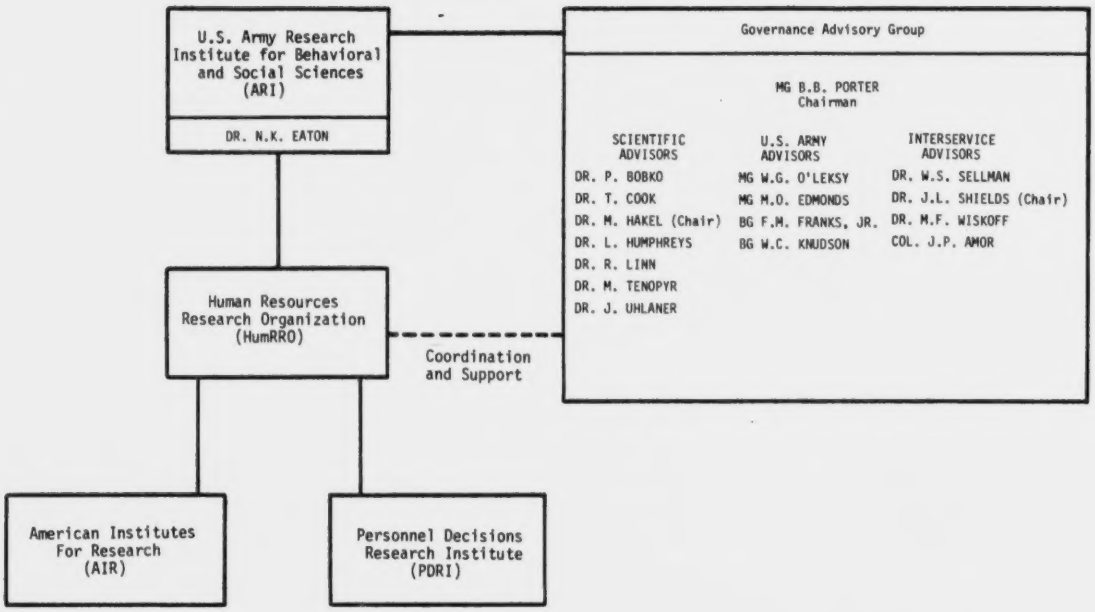


Figure 3. Governance Advisory Group.

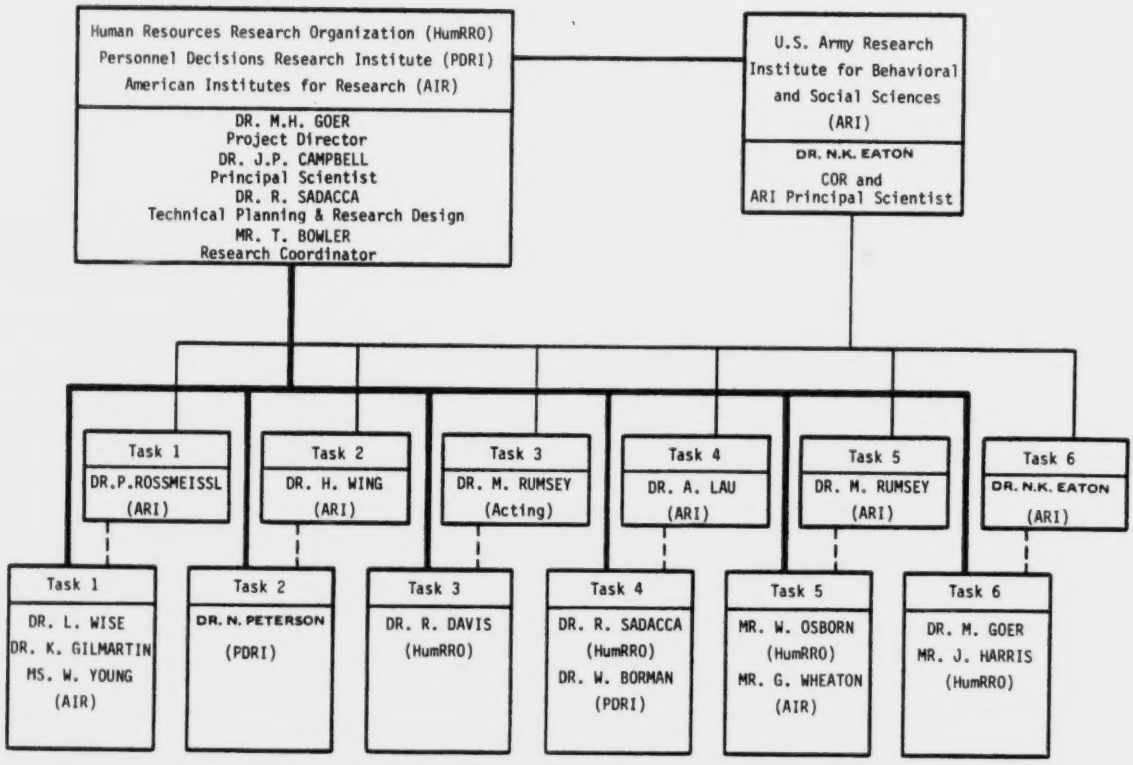


Figure 4. Project organization.

Documentation

The following relevant and related research reports and papers (see abstracts in Appendix A) were prepared during the 1984 fiscal year:

"Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report," by the Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute, ARI Research Report 1347.

"Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Technical Appendix to the Annual Report," Newell K. Eaton and Marvin H. Goer (Editors), ARI Research Note 83-37.

"Development and Validation of Army Selection and Classification Measures, Project A: Longitudinal Research Database Plan," by Lauress L. Wise, Ming-mei Wang, and Paul G. Rossmeissl, ARI Research Report 1356.

"The U.S. Army Research Project to Improve Selection and Classification Decisions," by Newell K. Eaton.

II. SCHOOL AND JOB PERFORMANCE MEASUREMENT

The overall objective for criterion measurement within Project A is to develop a broad array of valid and reliable criterion measures that reflect all major factors of job performance for first-tour enlisted personnel. These should constitute state-of-the-art criteria against which selection and classification measures can be validated.

Within this general objective the more specific purposes are to (a) determine the relationship of training performance to on-the-job performance, (b) measure performance "hands-on" by standardized simulations and work samples, and (c) compare rating scales, knowledge tests, and standardized work samples as alternative measures of specific task performance.

Project A is being conducted on a carefully selected sample of 19 MOS, as previously described. Using large samples of individuals from each of these 19 MOS, a major concurrent validation will be conducted in 1985 and a longitudinal validation will begin in 1986. Criterion measures that are specific to a particular MOS are being developed in "batches." The first batch (designated A or X) includes four MOS, the second batch (B/Y) five MOS, and the third batch (Z) 10 MOS.

Objectives for FY84

As described in the FY83 annual reports, Project A criterion development was at the following point at the beginning of the project's second year, in October 1983:

- The critical incident procedure had been used with two workshops of officers to develop a first set of 22 dimensions of Army-wide rating scales, as well as an overall performance scale and a scale for rating the potential of an individual to be an effective NCO.

- The critical incident procedure had also been used to develop dimensions of technical performance for each of the four MOS in Batch A (13B, cannon crewman; 64C, motor transport operator; 71L, administrative specialist; 95B, military police).

- A painstaking process had been used to select the pool of 30 tasks in each Batch A MOS that would be subjected to hands-on and/or knowledge test measurement. After preparing job task descriptions, the staff used a series of judgments by subject matter experts (SME), considering task importance, task difficulty, and intertask similarity, as the basis for selecting the final sets of tasks.

- On the way to developing norm-referenced training achievement tests for each of the 19 MOS, the staff had visited each proponent school and developed a description of the objectives and content of the training curriculum. They had also used Army Occupational Survey Program information to develop a detailed task description of job content for each MOS. After low-frequency elements were eliminated, SME judgments (N = 3-6) were used to rate the importance and error frequency for each task element. Approximately 225 tasks were then sampled proportionately from MOS duty areas. Consequently, at the end of FY83 we had a refined task sample for each MOS and systematic descriptions of the training program against which to develop a test item budget.

- A preliminary analysis had been made of the feasibility of obtaining archival performance records from the computerized Enlisted Master File (EMF) and the Official Military Personnel File (OMPF), which is centrally stored on microfiche. Because the OMPF data were incomplete, the staff decided to examine a sample of 201 Files (Military Personnel Records Jacket) to determine whether these files would be a more useful source of information.

The principal objectives for criterion development for FY84 were as follows:

(1) Use the information developed in FY83 to construct the initial version of each criterion measure.

(2) Pilot test each initial version and modify as appropriate.

(3) Evaluate the criterion measures for the four MOS in Batch A in a relatively large-scale field test (about 150 enlisted personnel in each MOS).

Construction of Initial Measures

Army-Wide Rating Scales. An additional four critical incident workshops involving 77 officers and NCOs were conducted during FY84. On the basis of the critical incidents collected in all workshops, a preliminary set of 15 Army-wide performance dimensions was identified and defined. Using a combination of workshop and mail survey participants (N = 61), the initial set of dimensions was retranslated and 11 Army-wide performance factors survived. The scaled critical incidents were used to define anchors for each scale, and directions and training materials for raters were developed and pretested.

During the same period scales were developed to rate overall performance and individual potential for success as an NCO. Finally, rating scales were constructed for each of 14 common tasks that were identified as part of the responsibility of each individual in every MOS.

MOS-Specific BARS Scales. Four critical incident workshops involving 70-75 officers and NCOs were completed for each of the MOS in Batch A and Batch B. A retranslation step similar to that for the Army-wide rating scales was carried out, and six to nine MOS-specific performance rating scales (Behaviorally Anchored Rating Scales, BARS) were developed for each MOS. Directions and training materials for scales were also developed and pretested.

Hands-On Measures (Batch A). After the 30 tasks per MOS were selected for Batch A, the two major development tasks that remained before actual preparation of tests were the review of the task lists by the proponent schools and the assignment of tasks to testing mode (i.e., hands-on job samples vs. knowledge testing).

The completeness and representativeness of the task lists were officially reviewed by the proponent school. Three of the reviews were conducted by mail and one through on-site briefing. Only slight changes were made in the task lists as a result of the reviews.

For assignment of tasks to testing mode, each task was rated by three to five project staff on three dimensions:

- The degree of physical skill required.
- The degree to which the task must be performed in a series of steps that cannot be omitted.
- The degree to which speed of performance is an important indicator of proficiency.

The extent to which a task was judged to require a high level of physical skill, a series of prescribed steps, and speed of performance determined whether it was assigned to the hands-on mode. For each MOS, 15 tasks were designated for hands-on measurement. Job knowledge test items were developed for all 30 tasks.

The pool of initial work samples for the hands-on measures was then generated from training manuals, field manuals, interviews with officers and job incumbents, and any other appropriate source. Each task "test" was designed to take from 5 to 10 minutes and was composed of a number of steps (e.g., in performing cardiopulmonary resuscitation), each of which was to be scored "go, no-go" by an incumbent NCO. A complete set of directions and training materials for scorers was developed; scorer training is thorough and is intended to take the better part of one day. The initial hands-on measures and scorer directions were then pretested on 5 to 10 job incumbents in each MOS and revised. They were ready for administration to the field test samples during the summer and fall of 1984.

MOS-Specific Job Knowledge Tests (Batch A). Concurrently, a paper-and-pencil, multiple-choice job knowledge test was developed to cover all of the 30 tasks in the MOS lists. The item content was generated on the basis of training materials, job analysis information, and interviews, with 4 to 10 items prepared for each of the 30 tasks. For the 15 tasks also measured

hands-on, the knowledge items were intended to be as parallel as possible to the steps that comprised the hands-on mode. The knowledge tests were pilot tested on approximately 10 job incumbents per MOS. After revision they were deemed ready for tryout with the field test samples.

Task Selection and Test Construction for Batch B. By the end of FY84, basic task descriptions had been developed for Batch B in a manner similar to that used for Batch A; that is, the CODAP (Comprehensive Occupational Data Analysis Program) and Soldier's Manual descriptions had been merged, edited to a uniform level of specificity, and evaluated for completeness and currency. The task descriptions have not yet been submitted to SME judgments of difficulty, importance, and similarity. The remaining steps of task selection, proponent review, assignment to testing mode, and test construction are scheduled for FY85.

In addition, for Batch B a formal experimental procedure is being used to determine the effects of scenario differences on SME judgment of task importance. The design calls for 30 SMEs to be randomly assigned to one of three scenarios (garrison duty/peacetime, full readiness for a European conflict, and an outbreak of hostilities in Europe). The implications of scenario differences are discussed later in this section.

Training Achievement Tests (Batch X). During FY84, generation of refined task lists for each of the 19 MOS in the Project A sample continued. For each MOS in Batch X (same MOS as Batch A), an item budget was prepared matching job duty areas to course content modules and specifying the number of items that should be written for each combination. An item pool that reflected the item budget was then written by a team of SMEs contracted for that purpose.

Next, training content SMEs and job content SMEs judged each item in terms of its importance for the job (under each of the three scenarios, in a repeated measures design), its relevance for training, and its difficulty. The items were then "retranslated" back into their respective duty areas by the job SMEs and into their respective training modules by the training SMEs. Items were designated as "job only" if they reflected task elements that were described as an important part of the job but had no match with training content; such items are intended to be a measure of incidental learning in training.

Once the sample of task elements was determined for each MOS and the items written and edited for basic clarity and relevance to the training, the job, or both, the pool was ready for tryout with the field test samples of incumbents and a sample of 50 trainers from each MOS.

Administrative (Archival) Indices. A major effort in FY84 was a systematic comparison of information found in the Enlisted Master File (EMF), the Official Military Personnel File (OMPF), and the Military Personnel Records Jacket (201 File). A sample of 750 incumbents, stratified by MOS and by location, was selected and the files searched. For the 201 Files the research team made on-site visits and used a previously developed protocol to record the relevant information. A total of 14 items of information,

including awards, letters of commendation, and disciplinary actions, seemed, on the basis of their base rates and judged relevance, to have at least some potential for service as criterion measures.

Unfortunately, the microfiche records appeared too incomplete to be useful and search of the 201 Files was cumbersome and expensive. It was decided to try out a self-report measure for the 14 administrative indices and compare it to actual 201 File information for the people in the field trials.

Batch A(X) Field Tests

The goal for the FY84 criterion field tests was to obtain enough information to permit relatively stable estimates of item and scale statistics, reliability indices, and scale/test intercorrelations. On the basis of these data, the array of criterion measures must be reduced to fit the time available (16 hours for Batch A/X and Batch B/Y MOS) for the FY83/84 concurrent validation sample which will be tested during the summer of 1985. The reduction must be accomplished by eliminating items and scales with psychometric deficiencies that cannot be fixed, redundant measures, and (if necessary) the least crucial parts of the criterion space.

Field Test Criterion Battery. The complete array of specific criterion measures that was actually used at each field test site is given below. For each rating scale every effort was made to obtain a complete set of supervisor, peer, and self ratings. This may very well be the most comprehensive array of performance measures ever used in a personnel research project.

A. MOS-Specific Performance Measures

1) Paper-and-pencil tests of knowledge of task procedures consisting of 4-10 items for each of 30 major job tasks for each MOS. Item scores can be aggregated in at least the following ways:
   - Sum of item scores for each of the 30 tasks.
   - Sum of item scores for common tasks.
   - Sum of item scores for MOS unique tasks.
   - Sum of item scores for 15 tasks also measured hands-on.

2) Hands-on measures of 15 tasks for each MOS.
   - Individual task scores.
   - Total score for common tasks.
   - Total score for unique tasks.

3) Ratings of performance on each of the 15 tasks measured via hands-on methods by:
   - Supervisors
   - Peers
   - Self

4) Behaviorally anchored rating scales of 5-9 performance dimensions for each MOS by:
   - Supervisors
   - Peers
   - Self

5) A general rating of overall job performance by:
   - Supervisors
   - Peers
   - Self

B. Army-Wide Measures

1) Eleven behaviorally anchored rating scales designed to assess the following dimensions. Three sets of ratings (i.e., from supervisors, peers, and self) were obtained on each scale for each individual.

   a) Technical Knowledge/Skill
   b) Initiative/Effort
   c) Following Regulations/Orders
   d) Integrity
   e) Leading and Supporting
   f) Maintaining Assigned Equipment
   g) Maintaining Living/Work Areas
   h) Military Appearance
   i) Physical Fitness
   j) Self-Development
   k) Self-Control

2) A rating of general overall effectiveness as a soldier by:
   - Supervisors
   - Peers
   - Self

3) A rating of NCO potential by:
   - Supervisors
   - Peers
   - Self

4) A rating of performance on each of 14 common tasks from the manual of common tasks by:
   - Supervisors
   - Peers
   - Self

5) A 14-item self-report measure of certain administrative indices such as awards, letters of commendation, and reenlistment eligibility.

6) The same administrative indices taken from 201 Files.

7) Attrit/not attrit during the first 180 days.

The Field Test Samples. The field test data were collected at different sites over a period of four months. Data for administrative specialists and military police were collected in U.S. installations during May, July, and August of 1984. Data on cannon crewmen and motor transport operators were obtained from two sites in Germany during August and September of 1984. The breakdown of subjects by MOS and by location is shown in Table 1. All subjects were incumbent enlisted personnel who had been in the Army 12 to 24 months.

Table 1. "Batch A" Field Test Samples

MOS		N
Administrative Specialists (71L)		129
Fort Polk	60	
Fort Hood	48	
Fort Riley	21	
Military Police (95B)		113
Fort Polk	42	
Fort Hood	42	
Fort Riley	29	
Cannon Crewmen (13B)		
Herzobase		150
Motor Transport Operators (64C)		
Mannheim		155
Total		547

Procedure. Staff members worked closely with the point of contact to secure testing sites, assemble equipment, and gain the cooperation of support personnel. The week before data collection, a project team visited the site to make sure everything was ready and to train the scorers of the hands-on measures. The tests and rating scales were administered by project personnel. Each participant was tested on each measure during a 2-day testing period. Approximately half the participants returned 6-12 days later and were retested on the hands-on measures. Every effort was made to obtain at least two supervisors and two peers to serve as raters for each incumbent on the rating scale measures. However, only one scorer was used for each hands-on task and scorers differed across tasks.

Analyses: Field Test Data. By the end of FY84, the field tests had been completed but the analyses of the data had not yet begun. To proceed from the current array of criterion measures to the set of measures to be used in the FY83/84 concurrent validation during 1985, a "Criterion Measures

Task Force" composed of appropriate consortium and ARI scientists and outside scientific advisers is being assembled. Their assignment is to systematically review the field test data and, through a series of decision meetings, eliminate poor quality or redundant measures, authorize revisions, and eventually make the reductions necessary to meet the concurrent validation time constraints. The first major meeting to review the field test data analysis was scheduled for November 1984.

Arriving at the criterion composites for the FY83/84 cohort validation is not the goal at this stage; those decisions will be a function of the FY83/84 concurrent validation data. The overall analysis objective is to reduce the amount of criterion measurement to fit the available time and at the same time maintain as broad a coverage of the criterion space as possible.

The specific objectives for the Criterion Measures Task Force are (a) to identify criterion measures that can be eliminated on the basis of poor psychometric quality or redundancy, and (b) to specify a prioritized list of options for reducing the Batch A criterion measures to fit the time constraints of the 1985 concurrent validation.

Confirmatory Analysis: A Beginning

After all analyses of the field test data are complete, Project A can take another step toward one of its major criterion development goals, the further refinement of the working model of soldier effectiveness. This could be done by first presenting the complete results of the field tests at a meeting of key task scientists and discussing them thoroughly. Next, task scientists would generate their own model of the criterion space. This would consist of naming and offering a definition for the latent variables, specifying how they are best measured by the available criteria, and describing any important features of the criterion space that he or she thinks are worth noting (e.g., "it is hierarchical in the following way ...").

Then a Delphi procedure could be used to show each model to everyone else and have each task produce a revised model. The revised models could be discussed at another group meeting to find out where there is agreement and disagreement about what the criterion space looks like. On the basis of that meeting, one or more alternative structural models that could be put to a confirmatory analysis in the FY83/84 cohort sample would be produced.

Discussion and Conclusions

As has been noted, the major accomplishments in criterion development for FY84 were:

(1) Construction, for four military jobs, of the initial operational versions of the largest and most comprehensive array of job performance criterion measures in the history of personnel selection/classification research.

(2) Revision and refinement of each measure through pilot testing.

(3) Development and pilot testing of training materials for raters and test administrators.

(4) Completion of a comprehensive field test of all criterion measures for four MOS, which involved two days of testing for approximately 600 job incumbents in several locations in the continental United States and in Europe.

(5) Preparation of the field test data for analysis.

Consequently, we now have the information necessary for making final revisions and for creating the final array of operational criterion measures for use for four MOS in the FY83/84 cohort concurrent validation during the summer of 1985. There is also an operational plan for how to analyze the field test data and an operational decisionmaking procedure for the final selection of criterion measures to be used in the concurrent validation.

During the past year a number of special issues have arisen that bear on criterion development in Project A. Some have been resolved and some are still under discussion. None have precise answers or are completely scientific in nature.

Scenario Effects. At several points in Project A, raters or SMEs are being asked to make judgments about such things as (a) the relative importance of specific job tasks to an MOS, (b) the relative importance of a knowledge test item for the objectives of a particular AIT program, (c) the degree of effective job performance reflected in a particular critical incident, (d) the job proficiency of a ratee on specific performance factors, and (e) the relative value (i.e., utility) of different job performance levels across MOS (e.g., How much more or less valuable to the Army is high performance for administrative specialists vs. low performance for motor transport operators?). It is often asserted that such judgments can be made meaningfully only when the context for the judgment (i.e., the scenario) is specified for the judge. For example, the relative importance of a specific task in the array of tasks that comprise an MOS can be judged only when the SME knows the context in which the task is to be performed (e.g., peacetime, wartime, field exercises).

There are two major reasons why differential scenario effects, if they exist, would be important for Project A.

First, they would influence the selection of content for all the criterion measures that we are using. For example, if job tasks vary in importance depending on the scenario, and hands-on or knowledge tests of task proficiency are to be constructed, then a wider variety of tasks may have to be included in the hands-on measure or knowledge test. That is, more items would be needed to cover all the important tasks if the subset of important tasks is not the same under each scenario.

Second, if the relative importance weights (i.e., utilities) for different MOS and for different performance levels within MOS vary substantially as a function of major scenario changes, then the selection/classification algorithm must incorporate different sets of utility weights which can be changed as the mission needs of the Army change.

To account for scenario differences in the selection of content for the MOS-specific job performance measures and the MOS-specific training performance measures, the following steps are currently being undertaken. For the five MOS in Batch B (same MOS as Batch Y), scenario effects on SME judgments of task importance are being studied experimentally. A total of 30 SMEs will be randomly assigned to one of three different scenarios, which are shown in Figure 5. Mean differences in importance ratings (by task and task cluster) will then be compared across scenarios.

The same three scenarios are being used in a repeated measures design to study scenario effects on judgments of item relevance for the knowledge tests to be used in Batch Y and Batch Z; SMEs are being asked to judge the relative importance of each knowledge test item for the content of the job. Each SME makes three importance judgments for each item corresponding to the three scenarios.

Results from the above steps will be used to determine whether scenario effects do in fact exist, and if so, for what types of tasks they are largest (e.g., common vs. MOS-specific). Preliminary results indicate that scenario effects on importance judgments are significant for certain kinds of tasks within some MOS. In particular, for non-combat support MOS the common tasks become more important and the MOS-specific tasks somewhat less important under a conflict rather than peacetime scenario.

Since some scenario effects do exist, the resolution has been to select tasks and test items that accommodate the differences. The preliminary data suggest that this should be possible within the constraints imposed by the FY83/84 concurrent validation design.

Multi-Method Measurement. In virtually any research project it is very desirable if the major variables can be measured by more than one method. In Project A, MOS-specific task performance is being assessed by three different methods (i.e., ratings, hands-on tests, and knowledge tests). Since testing time is not unlimited, a relevant issue is whether, for the concurrent validation, multiple measures should be retained at the expense of breadth of coverage, or vice versa. The relevant analyses that will inform this decision are not yet available, but the prevailing strategy is to do everything possible to preserve multiple measurement.

Weighting Criterion Components. Several measures in the criterion array are made up of component scores in the form of subtests on performance on complete tasks, as in the hands-on measures. A general issue concerns whether such components (e.g., the 15 separate hands-on tasks) should be differentially weighted before being combined into a total score. The same question arises when the aim is to combine specific criterion measures (e.g., ratings, knowledge tests, hands-on tests) into an overall composite for test validation.

1) Your unit is assigned to a U.S. Corps in Europe. Hostilities have broken out and the Corps' combat units are engaged. The Corps' mission is to defend, then reestablish, the host country's border. Pockets of enemy airborne/heliborne and guerilla elements are operating throughout the Corps sector area. The Corps maneuver terrain is rugged, hilly, and wooded, and weather is expected to be wet and cold. Limited initial and reactive chemical strikes have been employed but nuclear strikes have not been initiated. Air parity does exist.

2) Your unit is deployed to Europe as part of a U.S. Corps. The Corps' mission is to defend and maintain the host country's border during a period of escalating hostilities. The Corps maneuver terrain is inhibiting, weather is expected to be inclement. The enemy approximates a combined arms army and has nuclear and chemical capability. Air parity does exist. Enemy adheres to same environmental and tactical constraints as does U.S. Corps.

3) Your unit is a TO&E Field Artillery Battalion stationed on a military post in the Continental United States. The unit has personnel and equipment sufficient to make it mission capable for training and evaluation. The training cycle includes periodic field exercises, command and maintenance inspections, ARTEP evaluations, and individual soldier training/SQT testing. The unit participates in post installation responsibilities such as guard duty and grounds maintenance and provides personnel for ceremonies, burial details, and training support to other units.

Figure 5. Three alternative scenarios for SME judgments of task and item importance.

Two principal considerations govern the weighting of criterion components. First, the relative weight given to a particular component of job performance is a value judgment. Such judgments are part of the overall question of what an organization wants its people to be able to do. Weighting on other grounds, such as the relative reliability of measurement or degree of predictability, might produce composites in which the least important components are given the greatest weight. Second, the literature on differential weighting strongly suggests that if the number of components is very large (i.e., more than 4-6), then differential weighting makes very little difference in the psychometric properties of the total score.

Consequently, a reasonable strategy for Project A would be to compare weighted vs. unweighted criterion composites to determine whether differential weighting produces an advantage. The issue is scheduled to be considered during FY85.

Criterion Differences Across MOS. In Project A's validation of predictor measures for each of 19 jobs, the extent to which the same array of criterion measures will be used for the criterion composite in each MOS is a relevant question. For example, would job knowledge tests be used as a component of job performance in some MOS but not in others? This issue is being addressed directly by the continuing effort in Project A to develop an overall model of the effective soldier.

Within its current form, the model specifies the same set of constructs, or basic performance factors, for each MOS. In general, this means that very much the same measures would be used across MOS; however, their relative weights could vary considerably depending on the results of the MOS-specific development work and the criterion importance judgments. For example, the criterion factors assessed by the Army-wide rating scales could receive a much greater weight for combat MOS than for support MOS. Again, however, the most relevant data for informing this issue are not scheduled to be collected until FY85.

Potential Applications of FY84 Criterion Development Products

Since Project A is an R&D project designed to produce an improved selection and classification system for U.S. Army enlisted personnel, the purpose of criterion development is to produce optimal performance measures against which to validate new and improved selection and classification tests, rather than to produce new methods for operational performance appraisal. However, much of Project A's R&D work has operational implications. The major items that flow from the work during FY84 are as follows:

(1) The extensive work on the development of Army-wide performance factors via the critical incident workshops will provide a means both to confirm the validity of the current EER factors and to refine and extend the content of the EER if the Army so desires.

(2) The results of the 201 File analysis would be a valuable aid in any future attempts to refine the use of 201 File information in making future promotion or reenlistment decisions.

Documentation

The following relevant and related research reports and papers (see abstracts in Appendix A) were prepared during the 1984 fiscal year:

"An Analysis of SQT Scores as a Function of Aptitude Area Composite Scores for Logistics MOS," by Paul G. Rossmeissl and Newell K. Eaton.

"Administrative Records as Effectiveness Criteria: An Alternative Approach," by Barry J. Riegelhaupt, Carolyn DeMeyer Harris, and Robert Sadacca.

"Factors Relating to Peer and Supervisor Ratings of Job Performance," by Walter C. Borman, Leonard A. White, and Ilene F. Gast.

"Relationships Between Scales on an Army Work Environment Questionnaire and Measures of Performance," by Darlene M. Olson, Walter C. Borman, Loriann Roberson, and Sharon R. Rose.

"The Cost-Effectiveness of Hands-on and Knowledge Measures," by William Osborn and R. Gene Hoffman.

"Personal Constructs, Performance Schema, and 'Folk Theories' of Subordinate Effectiveness: Explorations in an Army Officer Sample," by Walter C. Borman.

"Development of a Model of Soldier Effectiveness," by Walter C. Borman, Stephan J. Motowidlo, Sharon R. Rose, and Lawrence Hanser.

III. PREDICTOR MEASUREMENT

The major activities completed during the second year of Project A with respect to predictor measure development were:

(1) The definition and identification of the most promising predictor constructs.

(2) The administration and initial analysis of the Preliminary Battery.

(3) The development, tryout, and pilot testing of the first version of the Trial Battery, called the Pilot Trial Battery.

(4) The development and tryout of psychomotor/perceptual measures, using a microprocessor-driven testing device.

All of these activities were aimed primarily at developing the Trial Battery, which will be completed and administered to a large sample of soldiers in the third year of Project A in accordance with the concurrent validation research design. Figure 6 is a flow chart of the major activities devoted to predictor measurement on Project A and shows the relationships between these activities. The numbers on the figure correspond to the activities listed above. Each of these activities is described briefly.

Predictor Development

Construct Definition. The first activity, defining and identifying the most promising predictor constructs, was accomplished in large part by using experts to provide structured, quantified estimates of the empirical relationships of a large number of predictors to a set of Army job performance dimensions (the dimensions were defined by other Project A researchers). By pooling the judgments of 35 experienced personnel psychologists, we were able to more reliably identify the "best" measures to carry forward in Project A.

These estimates were combined with other information (from the literature review and Preliminary Battery analyses) and evaluated by consortium and ARI scientists and members of the Scientific Advisory Group (SAG). A final, prioritized list of constructs was identified.

This effort also produced a heuristic model, based on factor analyses of the experts' judgments, that organizes the predictor constructs and job performance dimensions into broader, more generalized classes and shows the estimated relationships between the two sets of classes. This effort is fully described in Wing, Peterson, and Hoffman (1984).

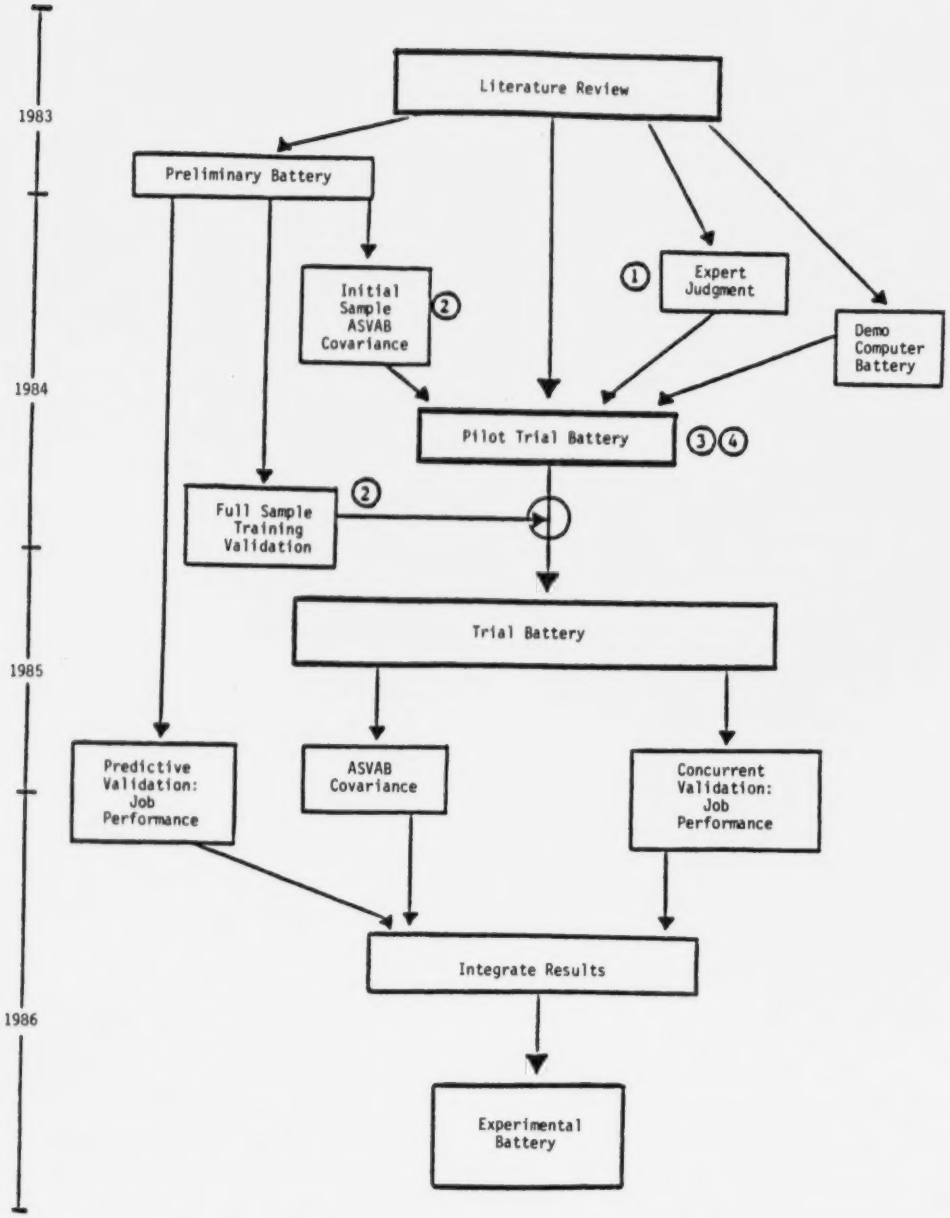


Figure 6. Flow chart of predictor measure development activities on Project A.

Preliminary Battery. Similarly, the initial analyses of Preliminary Battery data provided empirical results to guide our Pilot Trial Battery test development efforts. Data were collected with the Preliminary Battery on four MOS during the second year of the project. These four MOS were 05C (Fort Gordon), 19E/K (Fort Knox), 63B (Fort Dix and Fort Leonard Wood), and 71L (Fort Jackson).

The first 1800 cases from this sample were used in the initial analyses. These analyses enabled us to tailor the Pilot Trial Battery tests more closely to the enlisted soldier population. They also demonstrated the relative independence of cognitive ability tests and non-cognitive inventories of temperament, interest, and biographical data. This effort is fully reported in Hough et al. (1984).

A total of just over 11,000 Preliminary Battery cases were collected during Project A's second year. These data will be further analyzed to verify and extend the findings of the initial analyses. Most important, as Figure 6 indicates, the PB measures will be correlated with training performance measures to provide data for use in revising the Pilot Trial Battery during the third year of the project.

Pilot Trial Battery. The information from the first two activities fed into the third activity: the development, tryout, revision, and pilot testing of new predictor measures, collectively labeled the Pilot Trial Battery. New measures were developed to tap the ability constructs that had been identified and prioritized. These measures were tried out on three separate samples, with improvements being made between tryouts. The tryouts were conducted at Forts Carson, Campbell, and Lewis with approximately 225 soldiers participating.

At the end of the second year, the final version of the Pilot Trial Battery underwent a pilot test on a larger scale. Data were collected to allow investigation of various properties of the battery, including distribution characteristics, covariation with ASVAB tests, internal consistency and test-retest reliability, and susceptibility to faking and practice effects. About 650 soldiers participated in the pilot test.

Computerized Measures. The development, tryout, revision, and pilot testing of computerized measures is actually a subset of the Pilot Trial Battery development effort, but is worthy of separate mention. During the first year of the project, the literature review, site visits to military laboratories currently investigating computerized measures, and the programming of a demonstration battery laid the groundwork for FY84 activity.

Several objectives were reached during 1984. An appropriate microprocessor was identified and six copies were obtained for developmental use. The ability constructs to be measured were identified and prioritized. Software was written to utilize the microprocessor for measuring the abilities and to administer the new tests with an absolute minimum of human administrators' assistance. A customized response pedestal was designed and fabricated so that responses would be reliably and straightforwardly obtained from the people being tested. The software and hardware were put through an iterative tryout and revision process.

Pilot Trial Battery

Shown next is a general overview of the content of the Pilot Trial Battery, including the general ability area, method of measurement, number of tests or inventories, time to complete the tests, and total number of items.

Perceptual/Psychomotor Measures - Computer

Ten Tests
100 Minutes
343 Items

Cognitive Measures - Paper-and-Pencil

Ten Tests
100 Minutes
343 Items

Non-cognitive Measures - Paper-and-Pencil

Two Inventories
90 Minutes
Assessment of Background and Life Experiences (ABLE):
- Four Validity Scales
- Eleven Substantive Scales
- 270 Items

Army Vocational Interest Career Examination (AVOICE):
- Twenty-four Basic Interest Scales
- Six Organizational Climate/Environment Scales
- 309 Items

Figures 7 and 8 provide more detail about the substance of the Pilot Trial Battery. The cognitive/perceptual/psychomotor measures are shown in Figure 7. The predictor categories (left column) are the predictors that were identified as most promising, as described earlier. The Pilot Trial Battery test names are given in the right column. Note that ASVAB also appears in this column. This denotes that there is an ASVAB subtest that at least partially measures that predictor. Tests marked with an asterisk are administered via the computer-driven testing device.

Figure 8 shows the content of the two non-cognitive inventories, the Assessment of Background and Life Experiences (ABLE) and the Army Vocational Interest Career Examination (AVOICE). The AVOICE is a modified version of an inventory developed by the U.S. Air Force. Note that the Climate Environment Scales were not identified as essential predictors, but have been included at this point to measure individuals' perceptions of their organizations' environment.

Predictor Category	Pilot Trial Battery
Verbal	ASVAB
Memory	*Short Term Memory *Number Memory
Number Facility	ASVAB *Number Memory
Perceptual Speed and Accuracy	ASVAB *Perceptual Speed and Accuracy *Target Identification
Reasoning/Induction	Reasoning Test 1 Reasoning Test 2
Information Processing	*Simple Reaction Time *Choice Reaction Time
Spatial: Orientation	Orientation 1 Orientation 2 Orientation 3
Closure/Field Independence	Shapes
Spatial: Visualization	Object Rotations Assembling Objects Path Mazes
Mechanical Information	ASVAB
Multilimb Coordination	*Target Shoot *Target Tracking 2
Precision	*Target Shoot *Target Tracking 1
Movement Judgment	*Cannon Shoot

*Computerized

Figure 7. Cognitive/perceptual/psychomotor measures in the pilot trial battery.

Predictor Category	Pilot Trial Battery	
	AVOICE Scales	
Realistic vs. Artistic	Mechanics	Drafting
	Heavy Construction	Audiographics
	Marksman	Electronic Communication
	Electronics	Infantry
	Outdoors	Armor/Cannon
	Agriculture	Vehicle Operator
	Law Enforcement	Adventure
		Aesthetics
Investigative	Medical Service	
	Mathematics	
	Science/Chemical	
	Automated Data Procesing	
Enterprising Interests	Leadership	
Social Interaction	Teaching/Counseling	
Conventionality	Office Administration	
	Food Service	
	Supply Administration	
(N/A)	Climate Environment Scales	
	Achievement	Status
	Safety	Altruism
	Comfort	Autonomy
	ABLE Scales	
Stress Tolerance/Adjustment	Emotional Stability	
	Self-esteem	
Dependability/ Conscientiousness	Non-delinquency	
	Traditional Values	
	Conscientiousness	
Achievement/Work Orientation	Work Orientation	
Physical Condition/Athletic Abilities/Energy	Physical Condition	
	Energy Level	
Potency/Leadership	Dominance	
Locus of Control/ Work Orientation	Internal Control	
Agreeableness/Likability/ Sociability	Cooperativeness	

Figure 8. Non-cognitive measures in the pilot trial battery: The Army Vocational Interest and Career Examination (AVOICE) and the Assessment of Background and Life Experiences (ABLE).

Summation

At the end of the second year, the Pilot Trial Battery had been developed to measure a carefully identified and prioritized set of predictor constructs. It had been subjected to an iterative process of writing, trying out, and revising that resulted in a 6.5-hour battery of tests. Pilot test data were collected that will provide information for further refinement of the Pilot Trial Battery, especially a reduction in length. Ultimately this process will result in the Trial Battery that will be administered to over 12,000 soldiers in Year 3 of the project. In addition, more than 11,000 soldiers had completed the Preliminary Battery. Analyses of these data had informed the development of the Pilot Trial Battery, and further analyses will affect the refinement and reduction of the Pilot Trial Battery.

Documentation

The following relevant and related research reports (see abstracts in Appendix A) were prepared during the 1984 fiscal year:

"Validity of Cognitive Tests in Predicting Army Training Tests," by Clessen J. Martin, Paul G. Rossmeissl, and Hilda Wing.

"Expert Judgments of Predictor-Criterion Validity Relationships," by Hilda Wing, Norman G. Peterson, and R. Gene Hoffman.

"Covariance Analyses of Cognitive and Noncognitive Measures in Army Recruits: An Initial Sample of Preliminary Battery Data," by Leatta Hough, Marvin D. Dunnette, Hilda Wing, Janis Houston, and Norman G. Peterson.

"Meta-Analysis: Procedures, Practices, Pitfalls: Introductory Remarks," by Hilda Wing.

"Verbal Information Processing Paradigms: A Review of Theory and Methods," by Karen J. Mitchell, ARI Technical Report 648.

IV. VALIDATION

During Project A's second year, the Longitudinal Research Database (LRDB) was expanded dramatically to provide a firm basis for validation research. The first major validation research effort was carried out using information on existing predictors and criteria in the expanded LRDB. The initial validation research led to proposed improvements in the Army's existing procedures for selecting and classifying new recruits. The proposed improvements were adopted after thorough review and are to be implemented at the beginning of FY85. In addition, a number of smaller research efforts were supported with the expanded LRDB.

In describing validation research results during FY84, we turn first to an overview of the growth of the LRDB. Next, we summarize the ASVAB Aptitude Area Composite research that was based on the expanded LRDB. We conclude with a brief desription of other supporting analytic activities.

Growth of the LRDB

FY84 saw three major LRDB expansion activities. These were:

- The expansion of the FY81/82 cohort data files.
- The establishment of the FY83/84 cohort data files.
- The addition and processing of pilot and field test data files for different predictor and criterion instruments.

Each of these activities is described briefly.

Expansion of the FY81/82 Cohort Data Files. During FY83, we had accumulated application/accession information on all Army enlisted recruits who were processed in FY81 or FY82, and we had processed data from Advanced Instructional Training (AIT) courses on their success in training. During FY84, we added SQT data providing information on the first-tour performance of these soldiers subsequent to their training. SQT information was found for a total of 63,706 soldiers in this accession cohort, notwithstanding the fact that many of the soldiers in this cohort were not yet far enough along to be tested in this time period and others were in MOS which were not tested at all during this period.

In addition to SQT information, administrative information from the Army's Enlisted Master File (EMF) was added to the FY81/82 data base. Key among the variables culled from the EMF were those describing attrition from the Army, including the cause recorded for each attrition, and those describing the rate of progress of the remaining soldiers. Records were found for a total of 196,287 soldiers in this cohort. While the major source of administrative information was the FY83 year-end EMF files, information on progress and attrition was added from March and June 1984 quarterly EMF files.

Establishment of the FY83/84 Cohort Data Files. During FY84, application and accession information was assembled on recruits processed during FY83 and FY84. This cohort is of particular importance to Project A since it is the cohort to be tested in the concurrent validation effort. In addition to accession information, administrative data on the progress of this cohort also were extracted from annual and quarterly EMF files.

With the FY83/84 cohort, we began to include data collected on new instruments developed by Project A. Preliminary Test Battery information was collected on more than 11,000 soldiers in four different military occupational specialties. For three of these specialties (05C/31C, Radio/Teletype Operator; 71L, Administrative Clerk; and 63B, Light Wheel Vehicle Mechanic), data were collected at the beginning of AIT. In the fourth MOS (19E/K, Armor Crewman), data were collected at the beginning of combined Basic and AIT, generally within the first two weeks after accession. Data collected on these soldiers are described in Hough et al. (see Section III).

During FY84 we also collected data on success in AIT for soldiers in four MOS to which the Preliminary Battery was administered. At the end of FY84, data were still being added on soldiers who had taken the Preliminary Battery at the beginning of their training. The data collected included both written and hands-on performance measures administered at the end of individual modules as well as more comprehensive end-of-course measures. Table 2 shows the number of soldiers for whom Preliminary Battery information is available, the number of soldiers for whom training performance information is available, and the number of soldiers for whom both types of information are available.

Creation of Pilot and Field Test Data Files. During FY84, a great deal of information was collected in conjunction with the development of new instruments to be used in the FY85 concurrent validation. The largest accumulation of such information resulted from the Batch A combined criterion field test. (Batch A refers to the first four MOS of the nine MOS for which comprehensive performance measures are being developed.) In this effort, 548 soldiers in four different MOS each completed 2.5 days of testing. The tests administered included hands-on performance tests, job knowledge tests (both the task-specific version and the comprehensive tests being developed for use during training), and a wide range of rating data. (See Section II.) The combined information led to over 3,000 analysis variables for each of the soldiers tested.

A second major field test effort during FY84 was the Pilot Trial Battery field tests. These tests included both paper-and-pencil measures of aptitudes, interests, and background and the new computerized battery of perceptual and psychomotor tests. Scheduling conflicts postponed the data collection effort until the very end of the fiscal year, so initial processing of these data has only begun.

In addition to the major field tests of predictor and criterion instruments, data from a number of other efforts were incorporated into the LRDB. These included ratings of task and item importance, pilot tests on

Table 2. FY83/84 Soldiers with Preliminary Battery and Training Data

MOS	TOTAL PB CASES	TOTAL* TRAINING CASES	TOTAL CASES WITH BOTH PB & TRAINING DATA		
				%PB	%TR
05C/31C	2,411	1,971	833	(37)	(45)
19E/K	2,617	2,749	1,809	(69)	(66)
63B	3,245	1,959	1,223	(38)	(62)
71L	3,039	4,654	2,079	(68)	(45)
Total	11,312	11,313	5,944		

*As of FY84 year-end.

trainees of the comprehensive job knowledge tests intended for training use, and data gathered during the exploratory round of utility workshops.

ASVAB Area Composite Validation

As a first step in its continuing research effort to improve the Army's selection and classification system, Project A completed a large-scale investigation of the validity of Aptitude Area Composite tests used by the Army as standards for the selection and classification of enlisted personnel. This research had three major purposes: to use available data to determine the validity of the current operational composite system, to determine whether a four-composite system would work as well as the current nine-composite system, and to identify any potential improvements for the current system.

The Armed Services Vocational Aptitude Battery (ASVAB) is the primary instrument now used by the Armed Services for selecting and classifying enlisted personnel. The ASVAB is composed of ten cognitive tests or subtests, and these subtests are combined in various ways by each of the services to form Aptitude Area (AA) Composites. It is these AA composites that are used to predict an individual's expected performance in the service. The U.S. Army uses a system of nine AA composites to select and classify potential enlisted personnel: Clerical/Administrative (CL), Combat (CO), Electronics Repair (EL), Field Artillery (FA), General Maintenance (GM), Mechanical Maintenance (MM), Operators/Food (OF), Surveillance/Communications (SC), and Skilled Technical (ST).

The criterion measures used as indices of soldier performance in these analyses were end-of-course training grades and SQT scores. While both of these measures have some limitations, they were the best available measures of soldier performance. These two criterion measures were first standardized within MOS, and then combined to form a single index of a soldier's performance in his or her MOS.

One unique aspect of the composite development research was the large size of the samples used in the analyses. The sample sizes in the validity analyses for each of the AA composites are shown in Figure 9. The total sample size of nearly 65,000 soldiers renders this research one of the largest (if not the largest) validity investigations conducted to date.

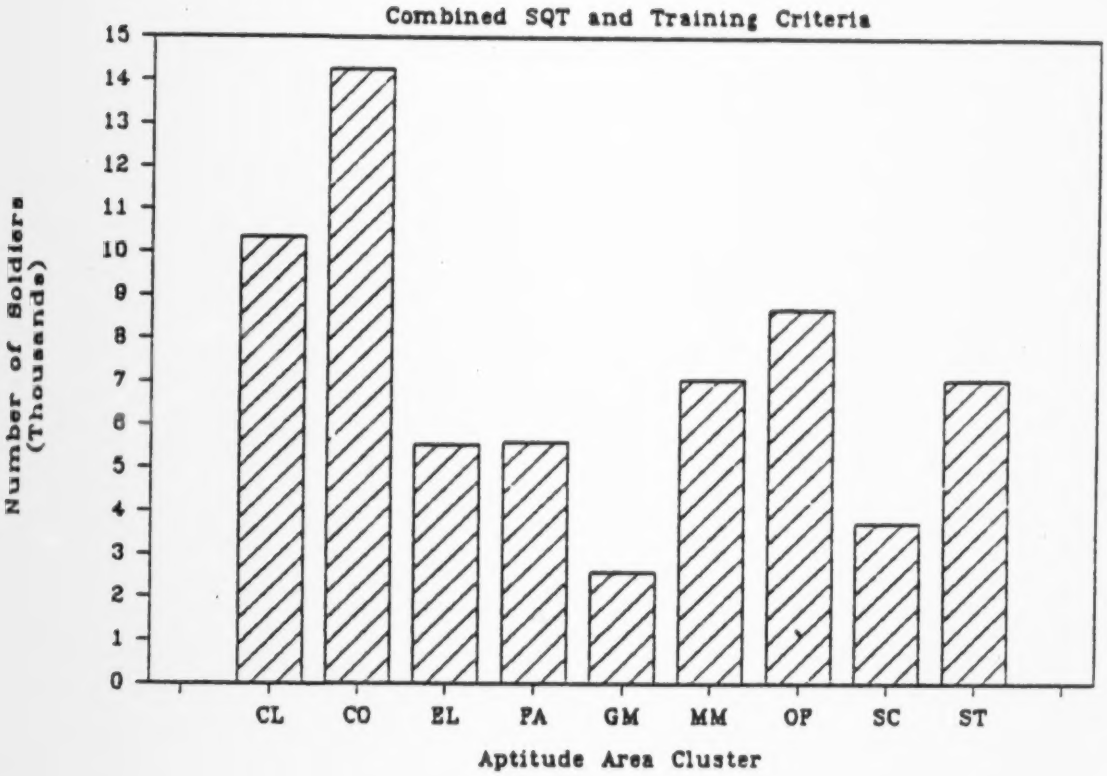


Figure 9. Validity analyses sample sizes.

The validities obtained in this research for the current nine AA composites are given in Figure 10. As can be seen, the existing composites are very good predictors of soldier performance. The composite validities ranged from a low of .44 to a high of .58, with the average validity being about .48. These numbers are high as test validities go.

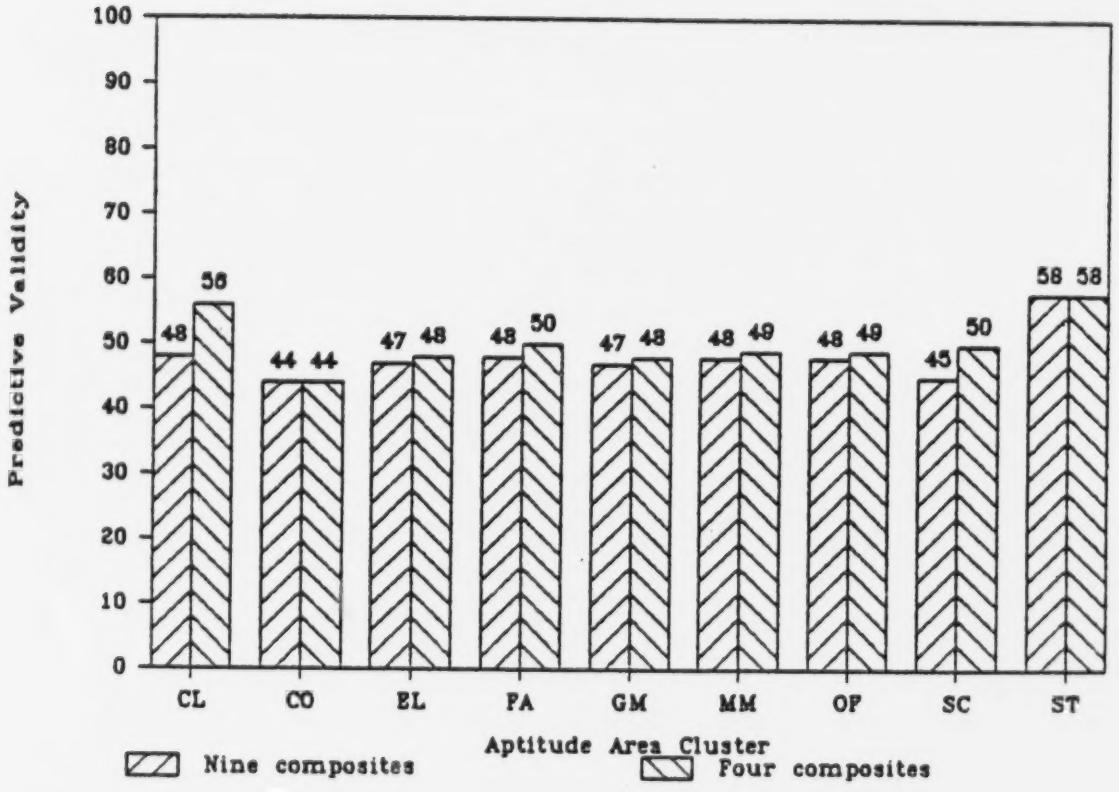


Figure 10. Predictive validities systems for nine and four composites.

A second finding of this research was that despite the high validities of the existing composites, a set of four newly defined AA composites could be used to replace the current nine without a decrease in composite validity. This set of four alternative composites included: a new composite for the CL cluster of MOS; a single new composite for the CO, EL, FA, and GM MOS clusters; a single new composite for the GM, MM, OF, and SC MOS clusters; and a new composite for the ST cluster of MOS.

Figure 10 also shows the test validities (corrected for range restriction) for this four-composite system when it is used to predict performance in the nine clusters of MOS defined by the current system. In all cases the four-composite solution showed test validities equal to or greater than the existing nine-composite case.

A corollary finding of the investigation into the four-composite solution was that the validities for two of the nine composites could be substantially improved without making major changes to the entire system. This improvement was accomplished by dropping two speeded subtests (numerical operations and coding speed) from the CL and SC composites and replacing them with the arithmetic reasoning and mathematical knowledge subtests for the CL composite and the arithmetic reasoning and mechanical comprehension subtests for the SC composite. Figure 11 compares the old and new forms for the CL and SC composites. This simple substitution of different subtests was able to improve the predictive validity of the CL composite by 16 percent and of the SC composite by 11 percent.

Based upon these data the Army has decided to implement the proposed alternative composites for CL and SC, effective 1 October 1984. Using the techniques developed by Hunter and Schmidt (1982) (which assume that an

	Current Composite		Proposed Composite	
Clerical/Administrative MOS	(VE+NO+CS)	.48	(VE+AR+MK)	.56
Surveillance/Communications MOS	(VE+NO+CS+AS)	.45	(VE+AR+MC+AS)	.50

Figure 11. A comparison of current and alternative composites.

individual's salary provides an approximation of that individual's worth to the organization), it can be estimated that these changes could lead to increased performance in the CL and SC MOS worth approximately $5 million per year. A fuller discussion of the research entailed in the development and validation of the AA composites can be found in McLaughlin, Rossmeissl, Wise, Brandt, and Wang (1984).

LRDB Support Activities

The expanded LRDB was also used in support of a number of other analytic activities. One such activity was the creation of an initial workfile containing Preliminary Battery data from tests administered through December 1983. Analyses based on this file were used to inform the development of the Trial Battery as well as to preview results for the Preliminary Battery.

EMF information being added to the LRDB was also used in support of ARI efforts to analyze the effects of alternative criteria for second-tour reenlistment eligibility.

A number of analysis files were provided to ARI staff in support of in-house research. These include a MAP data workfile, a Transportation School criterion data workfile, SQT information for addition to cohort files, and a workfile containing data from the Work Environment Questionnaire.

Documentation

The following relevant and related research reports (see abstracts in Appendix A) were prepared during the 1984 fiscal year:

"Evaluation of the ASVAB 8/9/10 Clerical Composite for Predicting Training School Performance," by Mary M. Weltin and Beverly A. Popelka, ARI Technical Report 594.

"Clustering Military Occupations in Defining Selection and Classification Composites," by Lauress L. Wise, Donald H. McLaughlin, Paul G. Rossmeissl, and David A. Brandt.

"Differential Validity of ASVAB for Job Classification," by Don McLaughlin.

"Complex Cross-Validation of the Validity of a Predictor Battery," by David Brandt, Don McLaughlin, Lauress Wise, and Paul Rossmeissl.

"Subgroup Variation in the Validity of Army Aptitude Area Composites," by Paul G. Rossmeissl and David A. Brandt.

"Validation of Current and Alternative ASVAB Area Composites, Based on Training and SQT Information on FY81 and FY82 Enlisted Accessions," by D.H. McLaughlin, P.G. Rossmeissl, L.L. Wise, D.A. Brandt, and Ming-mei Wang, ARI Technical Report 651.

"A Data Base System for Validation Research," by Paul G. Rossmeissl, Lauress L. Wise, and Ming-mei Wang.

"The Application of Meta-Analytic Techniques in Estimating Selection/ Classification Parameters," by Paul G. Rossmeissl and Brian M. Stern (to be published as an ARI Technical Report).

"Adjustments for the Effects of Range Restriction on Composite Validity," by David Brandt, Donald H. McLaughlin, Lauress L. Wise, and Paul G. Rossmeissl.

"Alternate Methods of Estimating the Dollar Value of Performance," by Newell K. Eaton, Hilda Wing, and Karen J. Mitchell.

V. STATUS AND FUTURE DIRECTIONS OF ARMY SELECTION AND CLASSIFICATION RESEARCH

In the first two years of operation, the Army's Project A has provided impressive examples of ways in which to address current research problems, social issues, and policy questions of interest to military selection and classification scientists and managers. Two years' research by 50 scientists on this project have produced many empirical findings and research designs that we hope will prove fruitful during the coming years of the project and highly applicable to future research and practice in human resource management.

The principal goal of the research being conducted in Project A is to significantly improve overall enlisted performance by means of more accurate selection and classification. Together, better predictor tests and performance assessment will substantially increase classification accuracy, which in turn will mean better performance by the Army in the field. Further, Project A research will develop a wide range of new measures of enlisted job performance and further explication of the meaning of job performance in the Army. Completion of the new system is also expected to reduce personnel costs significantly and provide the Army's personnel managers with a powerful tool for evaluation and control.

Overall, the system should improve the readiness of the Army, and the performance satisfaction and career opportunities of individual soldiers. We continue to believe that these gains will be achieved most efficiently through a single, integrated research and development effort. As to future trends, it seems likely that we will have a greater opportunity to make real contributions to the productivity of our military organizations in the coming decades than in any previous time in the history of selection and classification research. We now have a much improved research technology with which to address the multitude of questions surrounding the goal of placing the right individual in the right job, to benefit both the individual and the organization.

Criterion development during FY84 resulted in the following specific accomplishments:

(1) Construction of the initial versions of the largest and most comprehensive array of job performance criterion measures in the history of personnel selection/classification research.

(2) Revision and refinement of each measure through pilot testing.

(3) Development and pilot testing of training materials for raters and test administrators.

(4) Completion of a comprehensive field test of all criterion measures, which involved two days of testing for approximately 600 job incumbents in several locations in the continental United States and in Europe.

Consequently, we have the information necessary for making final revisions and for creating the final array of criterion measures that will be used in the concurrent validation of the FY83/84 cohort during the summer of 1985.

For predictor test development FY84 may have been the most important year of the project. It was the period during which the final decisions about what to measure were made, and the full array of tests was developed, including state-of-the-art computerized measures. More than 11,000 soldiers had completed the tests that comprised the Preliminary Battery. By the end of FY84, the Pilot Trial Battery had been developed to measure a carefully identified and prioritized set of predictor constructs. This battery had been subjected to an iteration process of item construction, initial pilot tryouts, and several revision phases that resulted in a 6.5-hour battery of tests painstakingly constructed to measure as complete an array of the most relevant variables as possible. Extensive pilot test data were then collected to provide information for further refinement of the Pilot Trial Battery, especially a reduction in length.

Ultimately this process will result in the Trial Battery that will be administered to more than 12,000 soldiers in Year 3 of the project. Taking into account the 11,000 soldiers tested with the Preliminary Battery, together these two selection test batteries probably constitute the most carefully scrutinized and broadest array of selection and classification tests ever used in selection and classification research.

Also in FY84, as a first step in its many-faceted effort to improve the Army's selection and classification system, Project A completed a large-scale examination of the validity of the Aptitude Area Composite tests used by the Army as standards for selecting and classifying enlisted personnel. On the basis of these data, the Army has decided to implement the proposed alternative composites for CL (clerical) and SC (Surveillance/Communications) MOS, effective 1 October 1984. It can be estimated that these changes could lead to improved CL and SC MOS performance worth $5 million per year to the Army.

Further comment is warranted about a number of special issues bearing on criterion development that have arisen in Project A. Some have been resolved and some are still under discussion. None have precise answers or are completely scientific in nature.

Scenario Effects. At several points in Project A, raters or SMEs are being asked to make judgments about such things as (a) the relative importance of specific job tasks to an MOS, (b) the relative importance of a knowledge test item for the objectives of a particular AIT program, (c) the degree of effective job performance reflected in a particular critical incident, (d) the job proficiency of a ratee on specific performance factors, and (e) the relative value (i.e., utility) of different job performance levels across MOS.

Preliminary results indicate that "scenario" effects on judgments of importance are significant for certain kinds of tasks within some MOS. In particular, for non-combat support MOS the common tasks become more important and the MOS-specific tasks somewhat less important under a conflict rather than peacetime scenario.

Since some context effects do exist, the resolution has been to select tasks and test items that accommodate the differences. The preliminary data suggest that this should be possible within the constraints imposed by the FY83/84 concurrent validation design.

Multi-Method Measurement. In virtually any research project, measuring the major variables by more than one method is very desirable. In Project A, MOS-specific task performance is being assessed by three different methods (i.e., ratings, hands-on tests, and knowledge tests). Since testing time is not unlimited, a relevant issue is whether, for the concurrent validation, multiple measures should be retained at the expense of breadth of coverage, or vice versa. The relevant analyses that will inform this decision are not yet available, but the prevailing strategy is to do everything feasible to preserve multiple measurement.

Weighting of Criterion Components. Several measures in the criterion array are made up of component scores in the form of individual rating scales, knowledge subtests, or performance on a complete but singular task, as in the hands-on measures. A general issue concerns whether such components (e.g., the 15 separate hands-on tasks) should be differentially weighted before being combined into a total score. The same question arises when the aim is to combine specific criterion measures (e.g., ratings, knowledge tests, hands-on tests) into an overall composite for test validation.

The strategy that Project A will pursue is to compare weighted vs. unweighted criterion composites and determine whether differential weighting produces an advantage. The issue is scheduled to be considered during FY85.

Criterion Differences Across MOS. In Project A's validation of predictor measures for each of 19 MOS, the extent to which the same array of criterion measures should be used for the criterion composite in each MOS is a relevant question. This issue is being addressed directly by the continuing effort in Project A to develop an overall model of the effective soldier. In its current form, the model specifies the same set of constructs, or basic performance factors, for each MOS. In general, this means that very much the same measures would be used across MOS; however, their relative weights could vary considerably depending on the results of the MOS-specific development work and the criterion importance judgments.

These issues include some of the most central problems in selection and classification research. Prospects appear to be good that efforts under way in Project A will make substantial contributions toward resolving these, and other, significant inquiries. Three factors support this view: the administrative efficiency of large and integrated programatic efforts; the comprehensive and interrelated consideration of all of the practical, social, legal, and policy questions directed toward making the optimal use of our soldiers; and the application of the most sophisticated technology available to explore a wide range of scientific problems that offer promising prospects for effective solutions.

REFERENCES

Eaton, Newell K. & Goer, Marvin H. (Eds.) Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Technical Appendix to the Annual Report. ARI Research Note 83-37, Army Research Institute. Alexandria, VA. October 1983.

Eaton, Newell K., Goer, Marvin H., Harris, James H., and Zook, Lola M. (Eds.). Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year. ARI Technical Report 660, Army Research Institute. Alexandria, VA. October 1984.

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute. Improving the Selection, Classification and Utilization of Army Enlisted Personnel. Project A: Research Plan. ARI Research Report 1332, Army Research Institute. Alexandria, VA. May 1983.

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute. Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. ARI Research Report 1347, Army Research Institute. Alexandria, VA. October 1983.

Human Resources Research Organization, American Institute for Research, Personnel Decisions Research Institute, and Army Research Institute. Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Appendices to Annual Report, 1984 Fiscal Year. ARI Research Note 85-14, Army Research Institute. Alexandria, VA. (In press.)

Hunter, J.E. & Schmidt, F.L. "Fitting People to Jobs: The Impact of Personnel Selection on National Productivity." In M.D. Dunnette and E.A. Fleishman (Eds.), Human Performance and Productivity: Human Capability Assessment. Hillsdale, N.J.: Lawrence Erlbaum. 1982. 231-284.

Peterson, N.G. & Bownas, D.A. "Skill, Task Structure, and Performance Acquisition." In M.D. Dunnette & E.A. Fleishman (Eds.), Human Performance and Productivity: Human Capability Assessment. Hillsdale, N.J.: Lawrence Erlbaum. 1982.

Schmidt, F.L., Hunter, J.E., McKenzie, R., & Muldrow, T. "The Impact of Valid Selection Procedures on Workforce Productivity." Journal of Applied Psychology, 64, 609-626, 1979.

APPENDIX A

ABSTRACTS FOR REPORTS, ARTICLES, AND PAPERS REFERENCED IN THE TEXT, FOR THE 1984 FISCAL YEAR

I. General

II. Performance Measurement

III. Predictor Measurement

IV. Validation

I. GENERAL

ARI Research Report 1347*
IMPROVING THE SELECTION, CLASSIFICATION AND UTILIZATION OF
ARMY ENLISTED PERSONNEL: ANNUAL REPORT
Human Resources Research Organization
American Institutes for Research
Personnel Decisions Research Institute
Army Research Institute
(October 1983)

This Research Report describes the research performed during the first year of a project to develop a complete personnel system for selecting and classifying all entry-level enlisted personnel. In general, the first year's activities have been taken up by an intensive period of detailed planning, briefing advisory groups, preparing initial troop requests, and beginning comprehensive predictor and criterion development that will be the basis for later validation work. A detailed description of the first year's work is contained in the Annual Report Technical Appendix, ARI Research Note 83-37.

---

* Available from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633. Order Document No. ADA141807.

ARI Research Note 83-37*
IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL: TECHNICAL APPENDIX TO THE ANNUAL REPORT
Newell K. Eaton and Marvin H. Goer (Editors)
(October 1983)

This Research Note describes in detail research performed during the first year of a project to develop a complete personnel system for selecting and classifying all entry-level personnel. Its purpose is to document, in the context of the annual report (ARI Research Report 1347), a variety of technical papers associated with the project. In general, the first year's activities have been taken up by an intensive period of detailed planning, briefing advisory groups, preparing initial troop requests, and beginning comprehensive predictor and criterion development that will be the basis for later validation work. Research reports associated with the work reported are included.

* Available from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633. Order Document No. ADA137117.

ARI Research Report 1356*
DEVELOPMENT AND VALIDATION OF ARMY
SELECTION AND CLASSIFICATION MEASURES
PROJECT A: LONGITUDINAL RESEARCH DATABASE PLAN*

Lauress L. Wise and Ming-mei Wang
(AIR)
Paul G. Rossmeissl
(ARI)

(December 1983)

This research report describes plans for the development of a major longitudinal research database. The objective of this database is to support the development and validation of new predictors of Army performance and also new measures of Army performance against which the new predictors can be validated. This report describes the anticipated contents of the database, editing procedures for assuring the accuracy of the data entered, storage and access procedures, documentation and dissemination procedures, and database security procedures.

---

* Available from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633. Order Document No. ADA143615. This document was included in the FY83 annual report (ARI Research Note 83-37) prior to publication as a Research Report.

THE U.S. ARMY RESEARCH PROJECT TO IMPROVE SELECTION AND CLASSIFICATION DECISIONS*

Newell K. Eaton
(ARI)

This paper provides an overview of the Army's Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel, and summarizes the results from the first 18 months of work. This major research effort will tie together the selection, classification, and job allocation of enlisted soldiers so that personnel decisions can be made to optimize performance and the utilization of individual abilities. Many activities are under way to improve predictor validity and performance measurement. Improved individual recruiting, performance, and retention are expected because the system will be designed to make the best match between the Army's needs and the individual's qualifications.

---

* Paper presented at the National Security Industrial Association Conference on Personnel and Training Factors in System Effectiveness, in Springfield, Virginia, May 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

II. PERFORMANCE MEASUREMENT

AN ANALYSIS OF SQT SCORES AS A FUNCTION OF APTITUDE AREA COMPOSITE SCORES FOR LOGISTICS MOS*

Paul G. Rossmeissl and Newell K. Eaton
(ARI)

To provide information useful in choosing the minimum Aptitude Area (AA) score that would permit enlistment in a Military Occupational Specialty (MOS), AA scores for soldiers in four quartermaster MOS were compared with their subsequent scores on the Skill Qualification Test (SQT) for their MOS. The four MOS were 76C (N=154), 76V (N=167), 76W (N=427), and 94B (N=3,536). Data were obtained for soldiers who entered the Army during FY81/82 and received SQT scores during the first two quarters of the 1983 test year. In general, SQT performance was higher for soldiers with higher AA scores; each 5-point increase in the AA score level was associated with higher SQT scores. SQT performance was quite high, with 80% or more of the soldiers passing in three of the four MOS. However, one-third or more of the soldiers in these MOS had AA scores within five points of the minimum score for entry into that MOS; hence a relatively modest increase in the AA minimum score for eligibility would have a relatively major effect in excluding applicants.

---

* Issued as Selection and Classification Technical Area Working Paper 84-12 (April 1984). Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

ADMINISTRATIVE RECORDS AS EFFECTIVENESS CRITERIA: AN ALTERNATIVE APPROACH*

Barry J. Riegelhaupt, Carolyn DeMeyer Harris, and Robert Sadacca
(HumRRO)

Attempts to measure individual job performance are meaningful only if the criterion accurately depicts effective job performance. Performance ratings rely on human judgment and hence are subjective in nature; objective indexes, on the other hand, tend to be incomplete or contaminated by outside factors (e.g., opportunity bias). This study explored the problems of using the administrative indexes that appear in Army personnel records in establishing criteria for soldier effectiveness. Records data were collected from the Military Personnel Record Jackets (MPRJ) for a random sample of 650 soldiers who had been in the Army between 14 and 27 months, divided among five widely diversified but populous MOS, at five different Army posts. From an original list of 38 variables, the following six were chosen after coding and analysis as potentially useful criteria of soldier effectiveness: Eligible to Reenlist, Has Received Letter/Certificate, Has Received Award, Has Had Military Training Courses, Has Received Article 15/FLAG Action, Promotion Rate (Grades Advanced/Year).

* Paper presented at the Annual Convention of the American Psychological Association in Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

FACTORS RELATING TO PEER AND SUPERVISOR RATINGS OF JOB PERFORMANCE*

Walter C. Borman
(PDRI)
Leonard A. White and Ilene F. Gast
(ARI)

While personnel ratings have long been widely used in evaluating job performance, not much is known about how such appraisals are made and how they relate to other means of measuring performance. Recently, research attention has been turned to achieving a better understanding of the appraisal process. Toward this end, in this study supervisor and peer ratings of first-term Army enlisted personnel were examined as a function of several factors that potentially influence these ratings. The elements considered in this research are (1) component job performance factors, (2) "good soldier" factors, (3) interpersonal relationship factors, and (4) job knowledge and skill factors. Peer and supervisor ratings were provided for 60 administrative specialists and 42 military police. Correlations between overall job performance ratings and ratings on each of the factors identified as a potential influence on ratings were examined. The results suggest that supervisor and peer ratings of overall job performance reflect more attention paid to individuals' performance on the job than to their standing on factors less directly relevant to performance. It is noted that interpretation of the finding must be limited because of the nature of the research approach and the small size of the sample.

---

* Paper presented at the Annual Convention of the American Psychological Association in Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

RELATIONSHIPS BETWEEN SCALES ON AN ARMY WORK ENVIRONMENT QUESTIONNAIRE AND MEASURES OF PERFORMANCE*

Darlene M. Olson
(ARI)
Walter C. Borman, Loriann Roberson, and Sharon R. Rose
(PDRI)

To identify and assess environmental and situational influences that affect job performance of first-tour soldiers, a 110-item Army Work Environment Questionnaire (AWEQ) was developed and given a preliminary tryout with 102 enlisted personnel. The research identified 14 job- and climate-related environmental factors that appear important within the Army work environment, and represented these dimensions in scale form in the AWEQ. Nine of these factors are considered "job content-related" and five "climate-related." The AWEQ was administered on a pilot basis to first-term soldiers in MOS 95B (Military Police) and MOS 71L (Administrative Specialist), and supervisory and peer ratings of overall soldier effectiveness were also obtained for these soldiers to provide performance indices for comparison with the AWEQ ratings. AWEQ results proved to be significantly related to supervisory ratings of job performance for six environmental scales (Training, Job-Relevant Authority, Work Assignment, Rewards/Recognition/Positive Feedback, Discipline, Job-Related Support) and to peer ratings of job performance for six scales (Physical Working Conditions, Job-Relevant Information, Changes in Job Procedures, Rewards/Recognition/Positive Feedback, Job-Related Support, Leader/Peer Role Models). Analyses of the preliminary results produced suggestions for revision, further development, and broad-scale testing of the AWEQ as a potential aid to evaluating the effect of Army environment on personal performance.

---

* Paper presented at the Annual Convention of the American Psychological Association in Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

THE COST-EFFECTIVENESS OF HANDS-ON AND KNOWLEDGE MEASURES*

William Osborn and R. Gene Hoffman
(HumRRO)

While hands-on tests of task performance are conceded to be the most valid measures of job proficiency, their cost (in time, personnel, and equipment) is often prohibitive. Knowledge tests are less costly but often do not correlate well with hands-on measures. In assessing proficiency in an Army job specialty in Project A, knowledge tests would provide greater task coverage but lower validity than hands-on tests; cost-effective decisions about the mix of measures that would provide the highest validity per unit of cost could be made if the relationships between the two types of measure were established for different types of tasks, and if the relative costs of the methods were known. This paper (1) discusses bases for estimating relative costs of hands-on and knowledge tests, (2) explores approaches to comparing the effectiveness of the two methods in measuring job proficiency in various types of tasks, and (3) discusses the effect on content validity of various combinations of methods. The major importance of the procedures being explored in Project A lies in the attempts to estimate relationships among tasks and test methods.

---

* Paper presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

PERSONAL CONSTRUCTS, PERFORMANCE SCHEMA, AND "FOLK THEORIES" OF SUBORDINATE EFFECTIVENESS: EXPLORATIONS IN AN ARMY OFFICER SAMPLE*

Walter C. Borman
(PDRI)

This research employs personal construct theory (Kelly, 1955) to explore the content of categories or schema that might be used in making work performance judgments. Twenty-five experienced U.S. Army officers, focusing on the job of non-commissioned officer (first-line supervisor), generated independently a total of 189 personal work constructs they believe differentiate between effective and ineffective NCOs. The officer subjects defined numerically each of their own 6-10 constructs by rating the similarity between each of these constructs and each of 49 reference performance, ability, and personal characteristics concepts. Correlations were computed between the subject-provided similarity ratings for each construct, and the 189 x 189 matrix was factor analyzed. Six interpretable content factors were identified (e.g., Technical Proficiency, Organization), with 124 of the 189 constructs from 23 of the 25 subjects loading substantially on these factors. Findings here suggest that a core set of concepts is widely employed by these officers as personal work constructs, but that different officers emphasize different combinations of this core set. Thus, substantial between-officer similarities and differences are evident. The personal constructs elicited from officer subjects are likened to performance schema and "folk theories" of job performance. Research is needed to assess the stability of these constructs over time and in different work contexts and to assess the impact of constructs on perceptions and evauations of job performance.

---

* Selection and Classification Technical Area Working Paper. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

DEVELOPMENT OF A MODEL OF SOLDIER EFFECTIVENESS*

Walter C. Borman
(PDRI)
Stephan J. Motowidlo
(Pennsylvania State University)
Sharon R. Rose
(PDRI)
Lawrence M. Hanser
(ARI)

This report introduces a conceptual model of individual effectiveness that extends beyond successful performance on specific job tasks to include elements of organizational commitment, socialization, and morale. The notion is that these broad constructs represent important criterion behaviors that contribute to an individual's worth to his or her organization and to its effectiveness. The idea of the model is applied to the "job" of enlisted soldier in the U.S. Army, and 15 dimensions springing from the conceptual model are named and defined.

Empirical research was then conducted to explore these effectiveness constructs. The report presents results of behavioral analysis research to develop dimensions of soldier effectiveness. Seventy-seven Army officers and NCOs in six workshops generated a total of 1315 behavioral examples of soldier effectiveness. Although by no means a formal test of the individual effectiveness model, the content of the examples generated shows similarities to elements of the model. Eleven dimensions emerged from behavioral analysis work and these results are discussed. Also discussed are advantages to taking a broader perspective on the performance criterion space in studying individual effectiveness, particularly in a military organization.

---

* Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; the appendices are issued separately in ARI Research Note 85-14. Order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

III. PREDICTOR MEASUREMENT

VALIDITY OF COGNITIVE TESTS IN PREDICTING ARMY TRAINING SUCCESS*

Clessen J. Martin, Paul G. Rossmeissl, and Hilda Wing
(ARI)

The purpose of this research was to determine the validity of Forms 8/9/10 (introduced in October 1980) of the Armed Services Vocational Aptitude Battery (ASVAB) in predicting success in training, in relation to both the Armed Forces Qualification Test (AFQT) and the ten Army Aptitude Area (AA) composites. Data on end-of-training grades during 1981 were collected for all MOS with 100 or more entrants per year, but research analyses were limited to 11 MOS having a sufficient variance in end-of-course grade (a training score standard deviation >5) to be useful in assessing predictor validities. For the Army AA composites, the overall corrected validity coefficient was .52 for Blacks and .62 for Whites. In the MOS where validities could be analyzed separately for gender subgroups, the average corrected validity coefficient was .61 for males and .58 for females. For the AFQT, the average validity across all 11 MOS was .64, which suggests that the Army composites examined in this research contribute relatively little to differential prediction of success in training. These results are not surprising in view of the limited focus of this study. Ongoing research with more MOS, using job performance as well as training criteria, is expected to provide more definitive information.

---

* Paper presented at the Psychonomics Society, San Diego, November 1983. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

EXPERT JUDGMENTS OF PREDICTOR-CRITERION VALIDITY RELATIONSHIPS*

Hilda Wing
(ARI)
Norman G. Peterson
(PDRI)
R. Gene Hoffman
(HumRRO)

As part of the Project A expansion of evaluation approaches in selecting and classifying Army enlisted personnel, a technical review of possible predictor and criterion measures was conducted. This consisted of collecting and analyzing expert judgments of the relationships to be expected between the most promising predictor constructs and various performance factors. Predictor variables (including cognitive, perceptual, psychomotor, biographical, vocational interest, and temperament) were identified in MOS-specific initial training, and in generalized Army effectiveness performance categories. The expert reviewers--35 industrial, measurement, or differential psychologists experienced in personnel selection--estimated the validity of each of 53 predictors against each of 72 criteria. Reliability, descriptive, and factor and cluster analyses were performed on the resulting judgments. Matrices were developed to display the mean estimated validity for each predictor-criterion combination, along with the standard deviation of this mean estimate across variables; available for comparison are summary tables of empirical criterion-related validity coefficients from prior research. The analyses indicated that experts can estimate the validity of a wide variety of predictor-criterion relationships with a high degree of reliability and at least reasonable accuracy; more definitive information on accuracy will be available as criterion-related validity research continues in Project A.

---

* Presented at the Annual Convention of the American Psychological Association in Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; the appendices are issued separately in ARI Research Note 85-14. Order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

COVARIANCE ANALYSES OF COGNITIVE AND NONCOGNITIVE MEASURES IN ARMY RECRUITS: AN INITIAL SAMPLE OF PRELIMINARY BATTERY DATA*

Leatta Hough and Marvin D. Dunnette
(PDRI)
Hilda Wing
(ARI)
Janis Houston and Norman G. Peterson
(PDRI)

Since World War II, the Army has based decisions about selection and classification of enlisted personnel upon cognitive abilities as predictors and upon training performance as the primary criterion. Under Project A these areas will be expanded to include noncognitive constructs of perceptual and psychomotor abilities, vocational interests, background, and temperament; existing predictor and criterion measures are being improved and new measures developed. This paper analyzes data from an initial sample, tested during the first two months of a nine-month data collection period, of soldiers (recruits) administered a Preliminary Battery (PB) of measures not previously included in the Armed Services Vocational Aptitude Battery (ASVAB). The PB included eight perceptual-cognitive measures; 18 vocational interest scales; 5 temperament scales; and a biographical questionnaire that could be scaled for male, female, or combined measures. Respondents were 2,286 soldiers in training in one of four selected MOS at one of five Army posts during October-November 1983. Results from the various item analyses, factor analyses, and other analyses are discussed, with especial reference to findings that will provide the basis for revisions of these measures in later Project A work.

---

* Paper presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

META-ANALYSIS: PROCEDURES, PRACTICES, PITFALLS: INTRODUCTORY REMARKS*

Hilda Wing
(ARI)

These introductory remarks for a symposium on meta-analysis, a process for combining the results of research from different studies, provide examples of the intricacies of trying to use this research analysis tool without full understanding of the hazards and potential power of the process.

* Presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

ARI Technical Report 648*
VERBAL INFORMATION PROCESSING PARADIGMS:
A REVIEW OF THEORY AND METHODS
Karen J. Mitchell

The theory and research methods of selected verbal information processing paradigms are reviewed. Work in factor analytic, information processing, chronometric analysis, componential analysis, and cognitive correlates psychology is discussed. The definition and measurement of cognitive processing operations, stores, and strategies involved in performance on verbal test items and test-like tasks is documented. Portions of the reviewed verbal processing paradigms are synthesized and a general model of text processing presented. The model was used as a conceptual framework for subsequent analyses of the construct and predictive validity of the verbal subtests of the Armed Services Vocational Aptitude Battery (ASVAB) 8/9/10.

---

* To be available from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633. This paper was included in the FY83 annual report (ARI Research Note 83-37) prior to publication as a Technical Report.

IV. VALIDATION

ARI Technical Report 594*
EVALUATION OF THE ASVAB 8/9/10 CLERICAL COMPOSITE
FOR PREDICTING TRAINING SCHOOL PERFORMANCE
Mary M. Weltin and Beverly A. Popelka
(October 1983)

The composite of Armed Services Vocational Aptitude Battery (ASVAB) subtests used to select applicants for entry-level training in Army clerical schools was evaluated by correlating composite scores with training performance scores. The clerical composite (CL) had high validity (r=.68) for this criterion, but an alternate composite of Arithmetic Reasoning, Paragraph Comprehension, and Mathematics Knowledge scores produced from multiple regression analyses had even higher validity (r=.74). Differential prediction for classification purposes is discussed.

---

* Available from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633. Order Document No. ADA143235. This paper was included in the FY83 annual report (ARI Research Note 83-37) prior to publication as a Technical Report.

CLUSTERING MILITARY OCCUPATIONS IN DEFINING SELECTION AND CLASSIFICATION COMPOSITES*

Lauress L. Wise and Donald H. McLaughlin
(AIR)
Paul G. Rossmeissl
(ARI)
David A. Brandt
(AIR)

The present Armed Services Vocational Aptitude Battery (ASVAB) is comprised of ten subtests, which are grouped in various combinations to identify and predict future performance in clusters of occupational specialties. Part of the Project A research is examining alternative clusterings of the entry-level Army MOS to define common predictor composites. This paper compares results from an initial investigation of use of several different clustering algorithms for ASVAB scores from recruits who entered the Army during FY81/82; subsequent selected Skill Qualification Test (SQT) results were used as the criterion measure. Because of lack of stability in the similarity measures, the attempt to cluster MOS on a purely empirical basis was abandoned, and work began on a system using a measure of loss of variance accounted for through substitution of the best unit weight composite for each cluster.

---

* Paper presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

DIFFERENTIAL VALIDITY OF ASVAB FOR JOB CLASSIFICATION*

Don McLaughlin
(AIR)

Since overall Army performance depends on how well recruit skills are matched to the requirements of the MOS the recruits enter, a set of ASVAB Aptitude Area composites must be evaluated in terms of its differential validity. The practical problem is that the best criterion for estimating differential validity is not available, since the same individual cannot be tested for performance in all jobs. This paper describes estimates for differential validity in (1) the case of unconstrained assignment, using a procedure devised by Horst (1954) to assess differential validity of a test battery, and (2) the case of constrained assignment, using a representative assignment algorithm. Alternative composites now under study indicated gains in comparison with the composites in current use.

---

* Paper presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

COMPLEX CROSS-VALIDATION OF THE VALIDITY OF A PREDICTOR BATTERY*

David Brandt, Don McLaughlin, and Laurie Wise
(AIR)
Paul Rossmeissl
(ARI)

This paper describes two uses of repeated replication methods to assess the stability of sample statistics in Armed Services Vocational Aptitude Battery (ASVAB) validation work. For the similarity matrix, an elementary repeated replication method (bootstrap) provided definitive answers. Sample statistics from two orthogonal replications correlated so poorly that further work on empirical clustering was abandoned. The bootstrap method produced estimates of errors that were reasonable when compared to classical error estimates of sample correlations. The standard errors for corrected validities were generally between one and two times the standard errors of the corresponding sample correlations. Especially large increases in standard errors were found in relatively small MOS with skewed distributions of criterion scores.

* Paper presented at the Annual Convention of the American Psychological Association in Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

SUBGROUP VARIATION IN THE VALIDITY OF ARMY APTITUDE AREA COMPOSITES*

Paul G. Rossmeissl
(ARI)
David A. Brandt
(AIR)

The current and proposed alternative Armed Services Vocational Aptitude Battery (ASVAB) Aptitude Area (AA) composites were investigated for possible subgroup bias in several ways. Analyses included predictive validities, comparisons of subgroup regression, lines, and plotting of the relationship of the subgroup regression and the common regression line. All subgroups were found to be well predicted by the composites. Both sets of composites showed small differences in predictive validity as a function of race and gender. The regression line comparisons indicate that, while some MOS (e.g., 76Y) need further research, in general either set of composites could be used to select and classify enlisted personnel for the Army without resulting in increased bias against blacks or women.

---

* Paper presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

ARI Technical Report 651*
VALIDATION OF CURRENT AND ALTERNATIVE ASVAB AREA COMPOSITES,
BASED ON TRAINING AND SQT INFORMATION ON
FY1981 AND FY1982 ENLISTED ACCESSIONS
D.H. McLaughlin, P.G. Rossmeissl, L.L. Wise,
D.A. Brandt, Ming-mei Wang

This report describes a large-scale research effort to validate and improve the Armed Services Vocational Aptitude Battery (ASVAB) Aptitude Area (AA) composites now used by the Army to select and classify enlisted personnel. Data were collected from existing Army sources on over 60,000 soldiers and over 60 MOS. The research had three major components: first, the composites now being used by the Army were validated; second, a new set of composites was derived empirically; finally, both sets were compared on the basis of predictive validity, differential validity, and possible prediction bias. Both sets of composites were found to perform well, with the alternative set of four composites doing slightly better than the nine now in operational use.

---

* To be available from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-763.

A DATA BASE SYSTEM FOR VALIDATION RESEARCH*

Paul G. Rossmeissl
(ARI)
Lauress L. Wise and Ming-mei Wang
(AIR)

Research progress under Project A over several years will depend heavily on a vast amount of interrelated data assembled to provide access to the many research teams involved and yet to protect the integrity and privacy of the data. The database management system selected was RAPID, a relational database system designed to accommodate large statistical data sets. RAPID provides a significant degree of data compression, convenient storage and access modes, and interfaces with other statistical packages, such as SAS and SPSS. Security of the database will be protected by routine encryption of soldier identity information, careful control of access to the database, and maintenance of log information. Procedures will be designed to balance the ease with which data can be accessed against the security of the database.

---

* Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

THE APPLICATION OF META-ANALYTIC TECHNIQUES IN ESTIMATING SELECTION/CLASSIFICATION PARAMETERS*

Paul G. Rossmeissl and Brian M. Stern
(ARI)

Exploring the long-standing problem of combining findings from several research settings, this paper applies meta-analytic techniques proposed by Hunter, Schmidt, and Jackson (1982) to the investigation of criterion-related validity of cognitive tests. The concept underlying the approach is that the variance of any statistic can be divided into components corresponding to true and error variance. These techniques were used to examine ASVAB test validities for 11 military occupational specialties (MOS), against an end-of-training score criterion. The uncorrected validities gave little indication that the cognitive tests could predict training performance. However, application of the meta-analysis corrections yielded estimated true validities that were quite high--.56 for the Armed Services Vocational Aptitude Battery (ASVAB) subtests and .65 for the Army composites. These results indicate that cognitive tests can be accurate predictors of training success and also illustrate the value of combining the subtests into composites.

---

* Paper presented at the Psychonomics Society, San Diego, November 1983. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

ADJUSTMENTS FOR THE EFFECTS OF RANGE RESTRICTION ON COMPOSITE VALIDITY*

David Brandt, Donald H. McLaughlin, and Lauress L. Wise
(AIR)
Paul G. Rossmeissl
(ARI)

This paper presents the adjusted validities of the nine Armed Services Vocational Aptitude Battery (ASVAB) composites currently in operational use by the Army in the selection and classification of enlisted personnel. The predictive validity coefficients indicate the extent to which the composites can cover the skills needed to become proficient in the corresponding MOS, as measured by training outcomes and SQT scores. The results from the various validity analyses indicate that, in general, the current composites provide information relevant to predicting performance in training and on the job. It was noted that performance was below average on the composite that included both of the speeded tests (CS and NO). Validity coefficients show little variability within a given MOS cluster, but there is little evidence that the composites capture skills specific to targeted MOS jobs.

---

* Paper presented at the Annual Convention of the American Psychological Association at Toronto, Canada, August 1984. Available as part of Eaton, N.K., Goer, M.H., Harris, J.H., and Zook, L.M. (Eds.), Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Annual Report, 1984 Fiscal Year, U.S. Army Research Institute Technical Report 660, Alexandria, VA, October 1984; order from Defense Technical Information Center, 5010 Duke Street, Alexandria, VA, 22314. Phone: (202) 274-7633.

ALTERNATE METHODS OF ESTIMATING THE DOLLAR VALUE OF PERFORMANCE*

Newell K. Eaton, Hilda Wing, and Karen J. Mitchell
(ARI)

The standard deviation of performance quality measured in dollars, SD$, is critical to calculating the utility of personnel decisions. In one popular technique for obtaining SD$, supervisors estimate the dollar value of performance at different levels. In many cases supervisors can base estimates on the cost of contracting out the various levels of performance. Estimation problems can arise, however, where contracting out is not possible, as in government organizations without private industry counterparts, or where individual salary is only a small percentage of the value of the performance to the organization or of the equipment operated. This paper presents two strategies ("superior equivalents" and "system effectiveness") for estimating the value of performance and determining SD$ by considering the changes in the numbers and performance levels of system units that lead to improved performance. One hundred Army tank commanders provided data about their jobs for these two strategies, as well as for the currently used "supervisor estimation" and "salary percentage" strategies. The new strategies appear to provide more appropriate and acceptable values of SD$ for those complex, expensive systems where dollar values of performance are less easily estimated.

* Personnel Psychology, 38, 27-40, 1985.